GREENSPAN'S BASIC & CLINICAL ENDOCRINOLOGY

David G. Gardner • Dolores Shoback

10th Edition

Greenspan's Basic & Clinical Endocrinology

Tenth Edition

Edited by

David G. Gardner, MD, MS

*Mount Zion Health Fund Distinguished Professor of
    Endocrinology and Medicine*
Chief, Division of Endocrinology and Metabolism
Department of Medicine and Diabetes Center
University of California, San Francisco

Dolores Shoback, MD

Professor of Medicine
Department of Medicine
University of California, San Francisco
*Staff Physician, Endocrine-Metabolism Section,
    Department of Medicine*
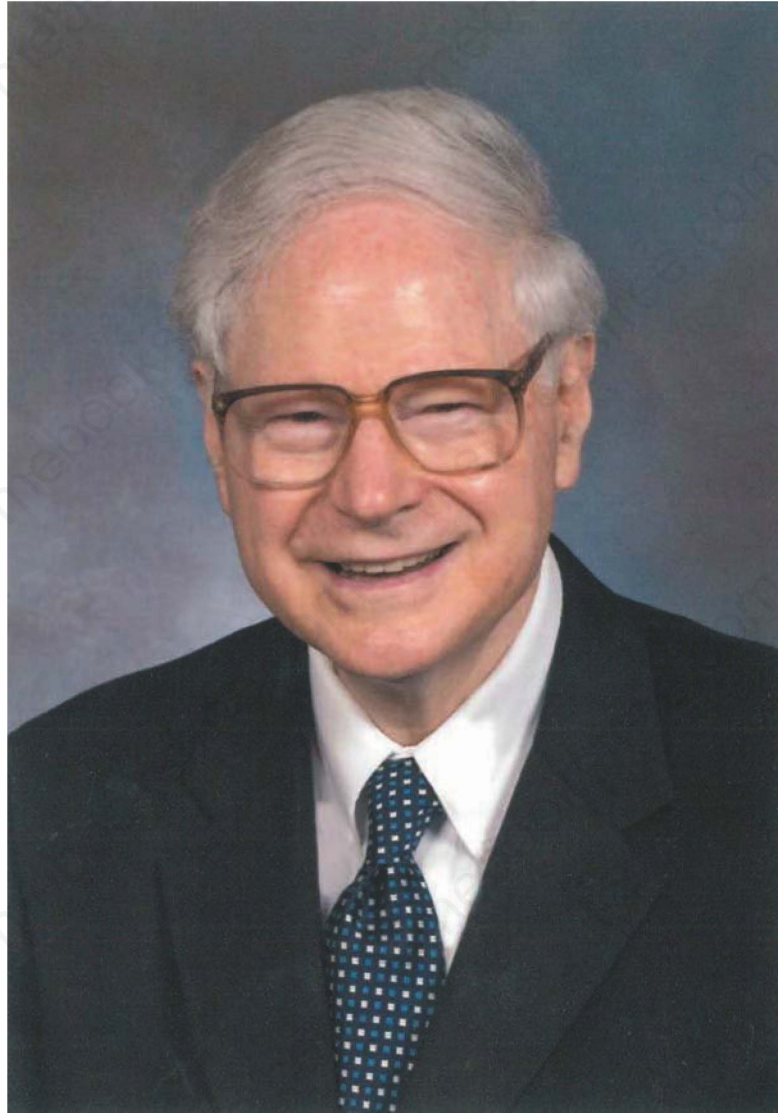*San Francisco Veterans Affairs Medical Center*

Notice

Medicine is an ever-changing science. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy are required. The authors and the publisher of this work have checked with sources believed to be reliable in their efforts to provide information that is complete and generally in accord with the standards accepted at the time of publication. However, in view of the possibility of human error or changes in medical sciences, neither the authors nor the publisher nor any other party who has been involved in the preparation or publication of this work warrants that the information contained herein is in every respect accurate or complete, and they disclaim all responsibility for any errors or omissions or for the results obtained from use of the information contained in this work. Readers are encouraged to confirm the information contained herein with other sources. For example and in particular, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this work is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs.

Francis Sorrel Greenspan, M.D. (1920-2016)

The tenth edition of *Greenspan's Basic & Clinical Endocrinology* is dedicated to the memories of four outstanding endocrinologists—Dr. John Baxter, Dr. Claude Arnaud, Dr. Melvin Grumbach, and, most especially, Dr. Francis Greenspan who was responsible for taking the initial steps to assemble this textbook more than thirty years ago. Each of these individuals was an outstanding endocrine scientist and/or clinical endocrinologist in the global endocrine community, and each contributed enormously to the success of this textbook.

Contents

5. The Posterior Pituitary (Neurohypophysis)   121

Alan G. Robinson, MD

6. Growth   137

Dennis Styne, MD

7. The Thyroid Gland   171

David S. Cooper, MD and
Paul W. Ladenson, MD (Oxon)., MD

8. Metabolic Bone Disease   239

Dolores M. Shoback, MD, Anne L. Schafer, MD, and Daniel D. Bikle, MD, PhD

9. Glucocorticoids and Adrenal Androgens 299

**Ty B. Carroll, MD, David C. Aron, MD, MS,
James W. Findling, MD, and J. Blake Tyrrell, MD**

10. Endocrine Hypertension   343

William F. Young, Jr, MD, MSc

11. Adrenal Medulla and Paraganglia   359

Paul A. Fitzgerald, MD

12. Testes   413

Bradley D. Anawalt, MD and
Glenn D. Braunstein, MD

13. Female Reproductive Endocrinology and Infertility   443

Mitchell P. Rosen, MD and Marcelle I. Cedars, MD

14. Disorders of Sex Development   501

Rodolfo A. Rey, MD, PhD,
Christopher P. Houk, MD,
Selma Witchel, MD, and Peter A. Lee, MD, PhD

15. Puberty   547

Dennis Styne, MD

16. The Endocrinology of Pregnancy   575

Bansari Patel, MD, Joshua F. Nitsche, MD, PhD, and Robert N. Taylor, MD, PhD

17. Pancreatic Hormones and Diabetes Mellitus   595

Umesh Masharani, MB, BS, MRCP (UK) and Michael S. German, MD

18. Hypoglycemic Disorders   683

**Umesh Masharani, MB, BS, MRCP (UK),
Stephen E. Gitelman, MD, and Roger K. Long, MD**

19. Disorders of Lipoprotein Metabolism   705

Mary J. Malloy, MD and John P. Kane, MD, PhD

20. Obesity   731

**Alka M. Kanaya, MD and
Christian Vaisse, MD, PhD**

21. Humoral Manifestations of Malignancy   743

Dolores M. Shoback, MD and Janet L. Funk, MD

22. Multiple Endocrine Neoplasia   757

David G. Gardner, MD, MS

23. Transgender Endocrinology   771

**Stephen M. Rosenthal, MD and
Wylie C. Hembree, MD**

24. Endocrine Emergencies   783

David G. Gardner, MD, MS

25. AIDS Endocrinopathies   809

Carl Grunfeld, MD, PhD

Authors

Bradley D. Anawalt, MD

Chief of Medicine
University of Washington Medical Center
Professor and Vice Chair
University of Washington Department of Medicine
Seattle, Washington
Testes

David C. Aron, MD, MS

Professor, Department of Medicine and Department of
 Epidemiology and Biostatistics, Division of Clinical and
 Molecular Endocrinology, School of Medicine, Case
 Western Reserve University; Associate Chief of Staff/
 Education, Louis Stokes Cleveland Department of Veterans
 Affairs Medical Center, Cleveland, Ohio
david.aron@med.va.gov
Evidence-Based Endocrinology and Clinical Epidemiology
Hypothalamus and Pituitary Gland
Glucocorticoids and Adrenal Androgens

Daniel D. Bikle, MD, PhD

Professor of Medicine and Dermatology, Veterans Affairs
 Medical Center and University of California, San Francisco
daniel.bikle@ucsf.edu
Metabolic Bone Disease

Glenn D. Braunstein, MD

Professor of Medicine
Cedars-Sinai Medical Center
Emeritus Professor of Medicine
The David Geffen School of Medicine at UCLA
Testes

Ty B. Carroll, MD

Assistant Professor, Endocrinology Center, Department of
 Medicine, Medical College of Wisconsin, Milwaukee
tcarroll@mcw.edu
Glucocorticoids and Adrenal Androgens

Marcelle I. Cedars, MD

Professor and Director, Division of Reproductive
 Endocrinology, Department of Obstetrics, Gynecology
 and Reproductive Sciences, University of California,
 San Francisco
marcelle.cedars@ucsfmedctr.org
Female Reproductive Endocrinology and Infertility

Orlo H. Clark, MD

Professor Emeritus of Surgery, Department of Surgery,
 University of California, San Francisco
clarko@surgery.ucsf.edu
Endocrine Surgery

David S. Cooper, MD

Professor of Medicine, Division of Endocrinology and
 Metabolism, Johns Hopkins University School of Medicine;
 Baltimore, Maryland
dscooper@jhmi.edu
The Thyroid Gland

James W. Findling, MD

Professor of Medicine, Director of Community Endocrine
 Services, Medical College of Wisconsin, Milwaukee
jfindling@mcw.edu
Hypothalamus and Pituitary Gland
Glucocorticoids and Adrenal Androgens

Paul A. Fitzgerald, MD

Clinical Professor of Medicine, Division of Endocrinology,
 Department of Medicine, University of California,
 San Francisco
paul.fitzgerald@ucsf.edu
Adrenal Medulla and Paraganglia

Janet L. Funk, MD

Associate Professor of Medicine, Division of Endocrinology,
 Department of Medicine, University of Arizona, Tucson
jfunk@u.arizona.edu
Humoral Manifestations of Malignancy

David G. Gardner, MD, MS

Mount Zion Health Fund Distinguished Professor of
 Endocrinology and Medicine; Chief, Division of
 Endocrinology and Metabolism, Department of Medicine
 and Diabetes Center, University of California, San Francisco
dgardner@diabetes.ucsf.edu
Hormones and Hormone Action
Multiple Endocrine Neoplasia
Endocrine Emergencies

Michael S. German, MD

Professor and Justine K. Schreyer Endowed Chair in Diabetes
 Research, Department of Medicine, Division of
 Endocrinology and Diabetes Center, University of
 California, San Francisco
mgerman@biochem.ucsf.edu
Pancreatic Hormones & Diabetes Mellitus

Stephen E. Gitelman, MD

Professor of Clinical Pediatrics, Chief, Division of Pediatric
 Endocrinology, Department of Pediatrics, University of
 California, San Francisco
sgitelma@peds.ucsf.edu
Hypoglycemic Disorders

Carl Grunfeld, MD, PhD

Professor of Medicine, University of California, San Francisco; Associate Chief of Staff for Research and Development; and Chief, Metabolism and Endocrine Sections, Veterans Affairs Medical Center, San Francisco
carl.grunfeld@ucsf.edu
AIDS Endocrinopathies

Wylie C. Hembree, MD

Associate Attending, New York Presbyterian Hospital; Retired Associate Professor of Medicine and of Obstetrics and Gynecology; Special Lecturer, Department of Medicine, Endocrine Division, College of Physicians and Surgeons, Columbia University Medical Center, New York, New York
wch2@columbia.edu
Transgender Endocrinology

Christopher P. Houk, MD

Associate Professor of Pediatrics; Chief, Pediatric Endocrinology, Medical College of Georgia, Georgia Regents University, Augusta, Georgia
chouk@gru.edu
Disorders of Sex Development

Edward C. Hsiao, MD, PhD

Associate Professor in Residence, Division of Endocrinology and Metabolism and Institute of Human Genetics, University of California, San Francisco
edward.hsiao@ucsf.edu
Hormones and Hormone Action

Juan Carlos Jaume, MD

Professor of Medicine; Chief, Division of Endocrinology, Diabetes and Metabolism; and Clinical Director of the Center for Diabetes and Endocrine Research (CeDER), College of Medicine and Life Sciences, University of Toledo, Toledo, Ohio
Juan.Jaume@utoledo.edu
Endocrine Autoimmunity

Bradley R. Javorsky, MD

Assistant Professor of Medicine, Endocrinology Center, Medical College of Wisconsin, Menomonee Falls
bjavorsky@mcw.edu
Hypothalamus and Pituitary Gland

Alka M. Kanaya, MD

Associate Professor of Medicine, Epidemiology & Biostatistics, University of California, San Francisco
alka.kanaya@ucsf.edu
Obesity

John P. Kane, MD, PhD

Professor Emeritus of Medicine, Biochemistry, and Biophysics, and Associate Director, Cardiovascular Research Institute, University of California, San Francisco
john.kane@ucsf.edu
Disorders of Lipoprotein Metabolism

Paul W. Ladenson, MD (Oxon)., MD

John Eager Howard Professor of Endocrinology and Metabolism; Professor of Medicine, Pathology, Oncology, and Radiology and Radiological Sciences; University Distinguished Professor, The Johns Hopkins University School of Medicine, Baltimore, Maryland
ladenson@jhmi.edu
The Thyroid Gland

Geeta Lal, MD, MSc, FRCS(C), FACS

Associate Professor of Surgery; Associate Chief Quality Officer, Adult Inpatient
University of Iowa, Iowa City, Iowa
geeta-lal@uiowa.edu
Endocrine Surgery

Peter A. Lee, MD, PhD

Professor of Pediatrics, Penn State College of Medicine, Hershey Medical Center, Hershey, Pennsylvania
plee@psu.edu
Disorders of Sex Development

Roger K. Long, MD

Associate Clinical Professor of Pediatrics, Division of Pediatric Endocrinology, University of California, San Francisco
Roger.Long@ucsf.edu
Hypoglycemic Disorders

Mary J. Malloy, MD

Professor (Emeritus), Department of Pediatrics and Medicine, Director, Pediatric Lipid Clinic and Co-Director, Adult Lipid Clinic, University of California, San Francisco
mary.malloy@ucsf.edu
Disorders of Lipoprotein Metabolism

Umesh Masharani, MB, BS, MRCP (UK)

Professor of Clinical Medicine, Division of Endocrinology and Metabolism, University of California, San Francisco
umesh.masharani@ucsf.edu
Pancreatic Hormones and Diabetes Mellitus
Hypoglycemic Disorders

Joshua F. Nitsche, MD, PhD

Assistant Professor, Department of Obstetrics and Gynecology, Wake Forest School of Medicine, Winston-Salem, North Carolina
jnitsche@wakehealth.edu
The Endocrinology of Pregnancy

Bansari Patel, MD

Assistant Professor, Wake Forest Baptist Medical Center,
Center for Reproductive Medicine, Winston-Salem,
North Carolina
bgpatel@wakehealth.edu
The Endocrinology of Pregnancy

Rodolfo A. Rey, MD, PhD

Director, Centro de Investigaciones Endocrinologicas
"Dr. Cesar Bergada", CONICET - FEI - Division de
Endocrinologia, Hospital de Ninos Ricardo Gutierrez,
Buenos Aires, Argentina
rodolforey@cedie.org.ar
Disorders of Sex Development

Alan G. Robinson, MD

Professor of Medicine, Associate Vice Chancellor, Medical
Sciences and Executive Associate Dean, David Geffen
School of Medicine at UCLA, University of California,
Los Angeles
robinson@ucla.edu
The Posterior Pituitary (Neurohypophysis)

Mitchell P. Rosen, MD

Associate Professor, Director, UCSF Fertility Preservation
Program and Reproductive Laboratories. Division of
Reproductive Endocrinology and Infertility, University of
California, San Francisco
Mitchell.Rosen@ucsf.edu
Female Reproductive Endocrinology and Infertility
Transgender Endocrinology

Stephen M. Rosenthal, MD

Professor Emeritus of Pediatrics, Division of Pediatric
Endocrinology; Medical Director, Child and Adolescent
Gender Center, University of California, San Francisco
stephen.rosenthal@ucsf.edu
Transgender Endocrinology

Anne L. Schafer, MD

Assistant Professor of Medicine, University of California, San
Francisco; Staff Physician, San Francisco Veterans Affairs
Medical Center, San Francisco, California
anne.schafer@ucsf.edu
Metabolic Bone Disease

Dolores M. Shoback, MD

Professor of Medicine, Department of Medicine, University of
California, San Francisco; Staff Physician, Endocrine-
Metabolism Section, Department of Medicine,
San Francisco Veterans Affairs Medical Center,
San Francisco, California
dolores.shoback@ucsf.edu
Metabolic Bone Disease
Humoral Manifestations of Malignancy

Ajay Sood, MD

Chief, Endocrinology Section, and Associate Professor of
Medicine, School of Medicine, Case Western Reserve
University and Louis Stokes Cleveland Department of
Veterans Affairs Medical Center, Cleveland, Ohio
ajay.sood@va.gov
Evidence-Based Endocrinology and Clinical Epidemiology

Dennis Styne, MD

Professor and Rumsey Chair, Department of Pediatrics,
Section of Endocrinology, University of California, Davis,
Sacramento
dmstyne@ucdavis.edu
Growth, Puberty

Robert N. Taylor, MD, PhD

Professor and Vice Chair for Research, Department of
Obstetrics and Gynecology; Co-Director, Molecular
Medicine and Translational Sciences Program, Wake Forest
School of Medicine, Winston-Salem, North Carolina
rtaylor@wakehealth.edu
The Endocrinology of Pregnancy

J. Blake Tyrrell, MD

Clinical Professor Emeritus of Medicine; Chief, Endocrine
Clinic, Division of Endocrinology and Metabolism,
University of California, San Francisco
blaket@medicine.ucsf.edu
Hypothalamus and Pituitary Gland
Glucocorticoids and Adrenal Androgens

Christian Vaisse, MD, PhD

Professor of Medicine, Department of Medicine,
Diabetes Center, University of California, San Francisco
vaisse@medicine.ucsf.edu
Obesity

Selma Witchel, MD

Director, Pediatric Endocrinology Fellowship Training
Program; and Associate Professor with Tenure, Children's
Hospital of Pittsburgh and University of Pittsburgh Medical
Center, Pittsburgh, Pennsylvania
witchelsf@upmc.edu
Disorders of Sex Development

William F. Young, Jr, MD, MSc

Professor of Medicine, Mayo Clinic College of Medicine,
Mayo Clinic, Rochester, Minnesota
young.william@mayo.edu
Endocrine Hypertension

Preface

This represents the tenth edition of *Greenspan's Basic & Clinical Endocrinology*—a bittersweet milestone in that it also marks the recent passing of Dr. Francis Greenspan, the originator and namesake of this textbook. Frank's involvement with this textbook will be sorely missed in the years to come. As with each of the previous editions, the individual chapters have been revised and updated to contain the most current information in the field. Our contributors continue to provide comprehensive content in a highly readable format. Chapter 14 (Disorders of Sex Development) has been completely revised and we have added a new chapter dealing with Transgender Endocrinology (Chapter 23). We trust that you have found previous versions of this text useful and informative and that the current version will continue to serve as a valuable tool for the education of your trainees and management of your endocrine patients.

David G. Gardner, MD, MS
Dolores Shoback, MD
San Francisco, CA

Hormones and Hormone Action

Edward C. Hsiao, MD, PhD and David G. Gardner, MD, MS

ACTH	Adrenocorticotropin hormone		**FAD**	Flavin adenine dinucleotide
ACVR1	Activin A receptor, type I		**FGF**	Fibroblast growth factor
AD1	Activation domain 1		**FMN**	Flavin mononucleotide
AD2	Activation domain 2		**FOX A1**	Forkhead transcription factor A1
AF-1	Activator function-1		**FXR**	Farnesoid X-activated receptor
AF-2	Activator function-2		**GAP**	GTPase-activating protein
Akt	Protein kinase B		**GAS**	Interferon gamma activated sequences
AMH	Anti-müllerian hormone		**GDP**	Guanosine diphosphate
ANP	Atrial natriuretic peptide		**GH**	Growth hormone
AP-1	Activator protein-1		**GHR**	Growth hormone receptor
APC	Adenomatous polyposis coli gene		**GLUT4**	Glucose transporter type 4
AR	Androgen receptor		**GR**	Glucocorticoid receptor
β-ARK	β-Adrenergic receptor kinase		**GRB2**	Growth factor receptor–bound protein-2
β-TrCP	Beta-transducin repeats-containing proteins		**GRE**	Glucocorticoid response element
BMP	Bone morphogenetic protein		**GRIP**	Glucocorticoid receptor–interacting protein
BNP	B-type natriuretic peptide		**GSK3**	Glycogen synthase kinase-3
BXR	Benzoate X receptor		**GTF**	General transcription factor
cAMP	Cyclic adenosine-3′,5′-monophosphate		**GTP**	Guanosine triphosphate
CAR	Constitutive androstane receptor		**HRE**	Hormone response element
CARM	Coactivator-associated arginine methyltransferase		**HSP**	Heat shock protein
CBP	CREB-binding protein		**ID**	Receptor–repressor interaction domain
cGMP	Cyclic guanosine-3′,5′-monophosphate		**IGF**	Insulin-like growth factor
CKI	Casein kinase I		**I-κB**	Inhibitor of nuclear factor kappa B
CNP	C-type natriuretic peptide		**IKK**	Inhibitor of nuclear factor kappa B kinase
CREB	cAMP response element-binding protein		IP_3	Inositol 1,4,5-trisphosphate
DAG	Diacylglycerol		IP_4	Inositol 1,3,4,5-tetrakis-phosphate
DAN	Differential screening-selected gene in neuroblastoma		**ISRE**	Interferon-stimulated response element
DBD	DNA-binding domain		**JAK**	Janus kinase
DRIP	Vitamin D receptor–interacting protein		**KHD**	Kinase homology domain
DVL	Dishevelled		**LBD**	Ligand-binding domain
EGF	Epidermal growth factor		**LH**	Luteinizing hormone
ER	Estrogen receptor		**LRP**	Lipoprotein receptor related protein
ERK	Extracellular signal–regulated kinase		**LXR**	Liver X receptor
			MAPK	Mitogen-activated protein kinase

MEK	MAPK kinase		**RAR**	Retinoic acid receptor
MR	Mineralocorticoid receptor		**RE**	Response element
MSH	Melanocyte-stimulating hormone		**RGS**	Regulators of G protein signaling
N-Cor	Nuclear receptor corepressor		**RSK**	Ribosomal S6 kinase
NF-κB	Nuclear factor kappa B		**RXR**	Retinoid X receptor
NO	Nitric oxide		**SH2**	*src* homology domain type 2
NOS	Nitric oxide synthase		**SIE**	Sis-inducible element
NPR	Natriuretic peptide receptor		**SMRT**	Silencing mediator for RXR and TR
NR	Nuclear receptor		**SOCS**	Suppressor of cytokine signaling
NRPTK	Non-receptor protein tyrosine kinase		**SOS**	Son-of-sevenless
PAK	p21-activated kinase		**SOST**	Sclerostin
P/CAF	p300/CBP-associated factor		**SR**	Steroid receptor
P/CIP	p300/CBP cointegrator-associated protein		**SRC**	Steroid receptor coactivator
PDE	Phosphodiesterase		**SRE**	Serum response element
PDGF	Platelet-derived growth factor		**SRF**	Serum response factor
PDK	Phosphatidylinositol-3,4,5 trisphosphate-dependent kinase		**STAT**	Signal transducer and activator of transcription
PHP-1a	Pseudohypoparathyroidism type 1a		**SWI/SNF**	ATP-dependent chromatin remodeling complex
PI-3K	Phosphoinositide-3-OH kinase		**TAZ**	WW domain-containing transcription regulator protein 1
PIP$_2$	Phosphatidylinositol-4,5-bisphosphate		**TBP**	TATA-binding protein
PIP$_3$	Phosphatidylinositol-3,4,5-trisphosphate		**TCF/LEF**	T-cell factor/lymphoid enhancer factor
PI(3,4)P2	Phosphatidylinositol-3,4-bisphosphate		**TGF-β**	Transforming growth factor beta
PKA	Protein kinase A		**TLE**	Transducin-like enhancer protein
PKB	Protein kinase B		**TPA**	12-*O*-tetradecanoyl-phorbol 13-acetate
PKC	Protein kinase C		**TR**	Thyroid hormone receptor
PKG	cGMP-dependent protein kinase		**TRAF**	Tumor necrosis factor receptor–associated factor
PLCβ	Phospholipase C beta		**TRAP**	Thyroid hormone receptor–associated protein
PLCγ	Phospholipase C gamma		**TRE**	TPA response element
PLC$_{PC}$	Phosphatidylcholine-selective phospholipase		**TSH**	Thyroid-stimulating hormone
POL II	RNA polymerase II		**VDR**	Vitamin D receptor
PPAR	Peroxisome proliferator–activated receptor		**Wnt**	int/Wingless family
PR	Progesterone receptor		**YAP**	Yes-associated protein-1
PTH	Parathyroid hormone			
PXR	Pregnane X receptor			
RANK	Receptor activator of nuclear factor kappa B			

Hormones are signaling molecules that traffic information from one point to another, typically through a soluble medium like the extracellular fluid or blood. Hormones fall into one of a number of different hormonal classes (eg, steroids, mono-amines, peptides, proteins, and eicosanoids) and signal through a variety of general (eg, nuclear vs cell surface) and specific (eg, tyrosine kinase vs phosphoinositide turnover) mechanisms in target cells.

Hormones produced in one tissue may promote activity in a target tissue at some distance from the point of secretion (endocrine effect). In this case the hormone travels through the bloodstream, often bound to a plasma protein, to access the target tissue. In addition, hormones may act locally following secretion; either on a neighboring cell (paracrine effect), on the secretory cell itself (autocrine effect), or without actually being released from the secretory cell (intracrine effect) (Figure 1–1).

Identification of a tissue as a target for a particular hormone requires the presence of receptors for the hormone in cells of the target tissue. These receptors, in turn, are linked to effector mechanisms that lead to the physiological effects associated with the hormone.

RELATIONSHIP TO THE NERVOUS SYSTEM

Many features of the endocrine system, such as the use of ligands and receptors to communicate between cells, are also found in the nervous system. In fact, from a functional standpoint, the two
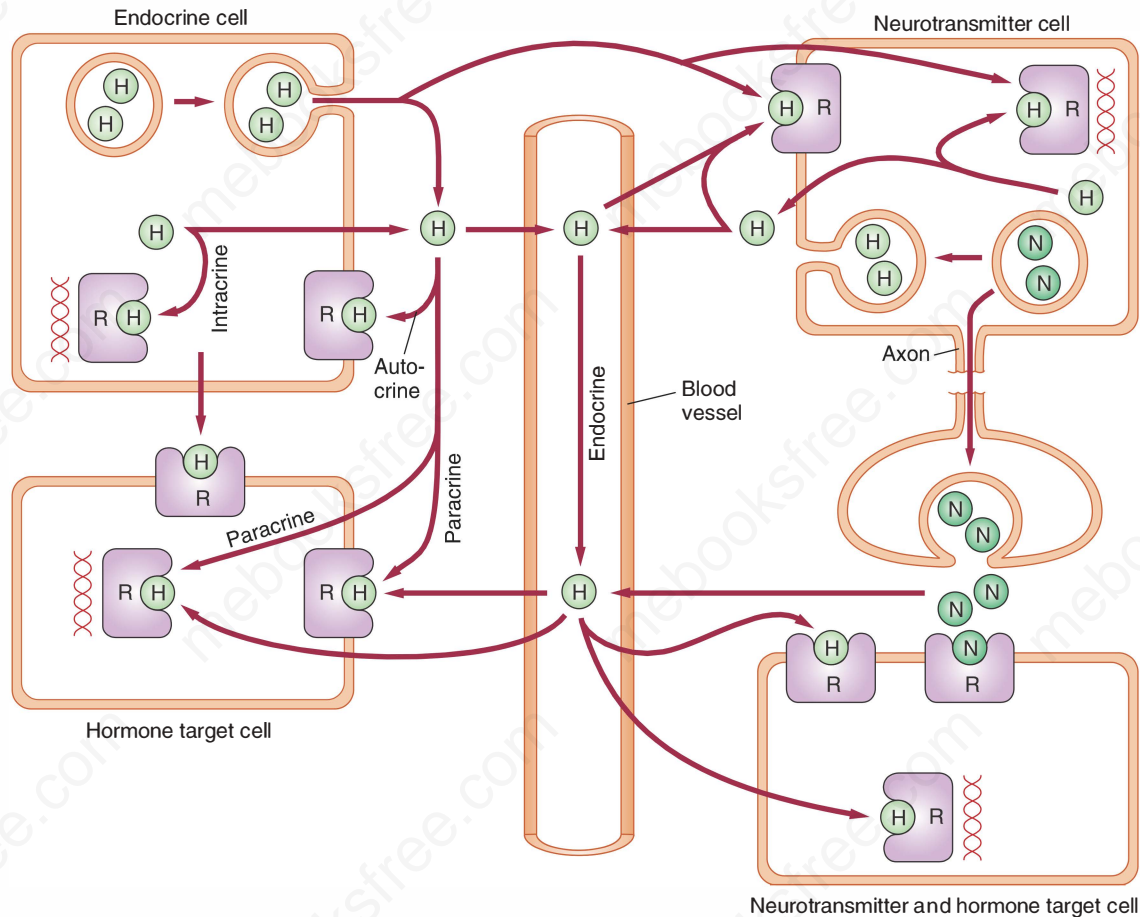
FIGURE 1–1 Actions of hormones and neurotransmitters. Endocrine and neurotransmitter cells synthesize hormones and release them by specialized secretory pathways or by diffusion. Hormones can act at the site of production either following release (autocrine) or without release (intracrine) from the producer cell. They can also act on neighboring target cells, including neurotransmitter-producing cells, without entering the circulation (paracrine). Finally, they can access target cells through the circulation (endocrine). Neurotransmitters that access the extracellular compartment, including circulating plasma, can act as paracrine or endocrine regulators of target cell activity (H, hormone; N, neurotransmitter; R, receptor).

systems are probably related evolutionarily. However, there are some important differences between the two systems. While the nervous system uses a highly compartmentalized, closed system of axons and dendrites to connect cells at some distance from one another, the endocrine system relies on circulating plasma to carry newly released hormones to their distant targets. As a result, the time constants for signal delivery are quite different between the two—virtually instantaneous for the nervous system but delayed, by virtue of circulation times, for the endocrine system. Thus, while neural responses are typically measured in seconds, endocrine responses are often measured in minutes to hours—thereby accommodating different needs in the organism. A second difference relates to the nature of the ligand–receptor interaction. In the nervous system, the affinity of receptor for ligand tends to be relatively low. This allows for rapid dissociation of ligand from receptor and, if that ligand is degraded locally, a rapid cessation of biological effect. Despite this rapid dissociation, the secretory neuron is able to maintain receptor occupancy by keeping concentrations of the ligand high around the target

neuron. It does this through pulsatile release of secretory granules into an incredibly small volume (ie, that determined by the volume in the synaptic cleft).

The endocrine system, on the other hand, has a very large volume of distribution for many of its ligands (eg, circulating blood volume). Maintaining ligand concentrations analogous to those present in the synaptic cleft would require prodigious secretory capacity. The endocrine system circumvents this problem by using ligand–receptor interactions with 100-10,000 fold higher binding affinity than those used in the nervous system. In effect, the nervous system is structured to deliver high ligand concentrations to relatively low-affinity receptors, allowing it to activate and inactivate biological effects quickly and in a relatively well-defined topography. Its effects are short lived. In contrast, the endocrine system uses high-affinity receptors to extract and retain ligand from a relatively "dilute" pool in circulating plasma. Its biological effects are long lasting. It has sacrificed rapid response to accommodate a wider area of signal distribution and prolongation of the biological effect. Thus, the systems are not only related but

complementary in the respective roles that they play in normal physiological function.

CHEMICAL NATURE OF HORMONES

Hormones vary widely in terms of their chemical composition. Specific examples include proteins (eg, adrenocorticotrophin), peptides (eg, vasopressin), monoamines (eg, norepinephrine), amino acid derivatives (eg, triiodothyronine), steroids (eg, cortisol), and lipids (eg, prostaglandins). Proteins can be glycosylated (eg, thyroid-stimulating hormone) and/or dimerized (eg, follicle-stimulating hormone) to generate full biological activity. In general, protein, peptide, monoamine, and lipophilic hormones tend to exert their effects primarily through protein receptors at the cell membrane, while thyroid hormone and steroids tend to operate in the cell nucleus. However, exceptions to these rules are being recognized (eg, triiodothyronine activates classic thyroid hormone receptors in the nuclear compartment and the trace amine receptor [TAR1] on the cell surface) and estradiol appears to activate both nuclear and plasma membrane receptors. It is likely that the biological "effect" of a given hormone reflects a composite of receptor activity located in several different cellular compartments.

ENDOCRINE GLANDS AND TARGET ORGANS

Endocrine glands are traditionally defined as ductless glandular structures that release their hormonal secretions into the extracellular space where they can eventually access circulating plasma. Classic endocrine glands include organs like the pituitary gland, thyroid gland, parathyroid glands, pancreatic islets, adrenal glands, ovaries, and testes. It is now clear that hormones can be secreted from nontraditional endocrine organs and play critical roles in the regulation of physiological homeostasis. Examples of the latter include the heart (natriuretic peptides), kidney (erythropoietin and renin), adipose tissue (leptin and adiponectin), bone (osteocalcin), and gut (cholecystokinin and incretins). Once in the circulation, hormones bind to receptors on target tissues to elicit their biological effects. Target tissues for some hormones (eg, glucocorticoids) are numerous, reflecting the ubiquitous distribution of their receptors, while those for other tissues have a more limited distribution (eg, androgens).

REGULATION OF HORMONE LEVELS IN PLASMA

Hormone levels in plasma determine the effective ligand concentration at the level of the hormone receptors in peripheral target cells. Thus, regulation of hormone levels plays an important role in the control of the biological effects that the hormone exerts.

Hormone Biosynthesis

New hormone synthesis is one of the principal mechanisms used to raise hormone levels in circulating plasma. In the case of protein or peptide hormones this usually reflects increased expression of the gene encoding the hormone (ie, increased production of the mRNA encoding the hormone) with subsequent increases in hormone synthesis. In the case of steroid or thyroid hormones it reflects increased sequestration of precursors for hormone synthesis (eg, cholesterol for steroid hormones or iodide for thyroid hormone) as well as increased activity of enzymatic proteins responsible for executing the individual catalytic events required for hormone production. The latter may involve a rate-limiting step in the synthetic cascade (eg, 1-alpha hydroxylase activity in the synthesis of 1,25-dihydroxyvitamin D).

Precursor Processing

Processing of hormone precursors contributes to varying degrees in controlling circulating hormone levels. Most peptide and protein hormones require some processing to generate the mature hormonal product (eg, conversion of proinsulin to insulin) and impairment in the processing activity can alter the ratio of precursor to product in plasma. In other cases, a critical processing event is part of the secretory process itself (eg, cleavage of thyroxine from thyroglobulin) and impaired processing can result in a dramatic reduction in immunoreactivity as well as bioactivity of the mature hormone. In addition, protein hormones may require post-translational modification (eg, glycosylation) or assembly (eg, heterodimerization) prior to secretion in order to optimize biological activity.

Hormone Release

Many hormones (eg, peptides, proteins, and monoamines) are stored in secretory granules in endocrine cells. Release of these granules is promoted by signaling events triggered by exogenous regulators termed secretagogues. This often requires activation of a second messenger system (see discussion under Receptors) like cyclic AMP generation or intracellular calcium mobilization in the endocrine cell. Steroid hormones, on the other hand, are not stored to a significant degree in the hormone-producing cells. In this case synthesis rather than hormone release appears to play the dominant role in controlling hormone levels in circulating plasma.

Hormone Binding in Plasma

Hormones in plasma can circulate either in a free form, uncomplexed with other molecules, or bound to other molecules like plasma proteins. It is the uncomplexed or free form of the hormone that represents the biologically active fraction of hormone in the plasma compartment, and it is this fraction which homeostatic regulatory mechanisms work to preserve.

However, binding of hormone to plasma proteins plays an important role in endocrine physiology. First, it provides a reservoir of hormone that exchanges with the free hormone fraction according to the laws of mass action (see under Receptors). This makes plasma hormone concentrations less dependent on hormone synthesis and release, effectively stabilizing those concentrations over extended periods of time. This also helps guarantee a uniform distribution of hormone concentration in capillary beds
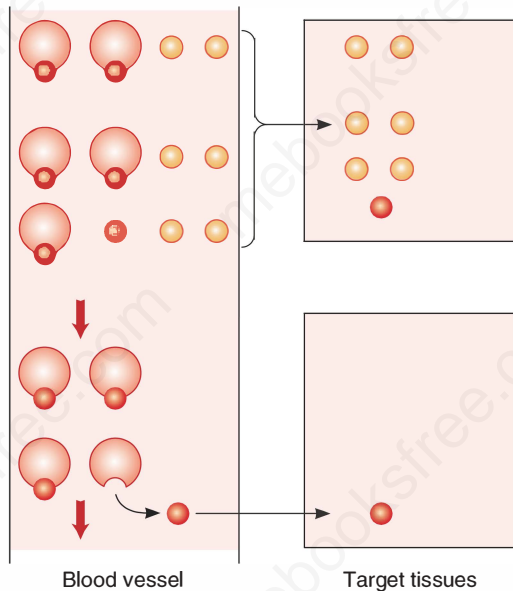
FIGURE 1–2 Role of plasma binding in delivery of hormones to peripheral tissues. Example shows a hormone that is bound (small red circles) to a plasma protein (large circles) and a hormone that is not protein bound (small orange circles). With the bound hormone, only the free fraction is available for tissue uptake. As the free fraction is depleted, additional hormone dissociates from the plasma-binding protein, making hormone available to more distal portions of the tissue. In contrast, all hormones that are not protein bound are quickly extracted in the proximal part of the tissue.

perfusing target tissues (Figure 1–2). Second, it slows the metabolism or turnover of the hormone by sequestering it away from degradative enzymes or filtration by the kidney.

Hormone Metabolism

Metabolism of hormones also plays an important role in regulating hormone concentrations. In some cases metabolism is responsible for converting precursors with less hormonal activity to products with greater activity (eg, conversion of 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D, or conversion of androstenedione to testosterone). In other cases, metabolism leads to degradation and inactivation of the hormone with a cessation of hormone activity. This type of degradation is often specific to the hormonal class under examination. Steroids, for example, are catalytically converted to inactive metabolites and/or sulfated to promote excretion. Thyroid hormones are subjected to deiodination which strips them of their biological activity. Protein and peptide hormones are internalized by target, as well as nontarget, cells and degraded in intracellular lysosomes. In general, the more avid the degradative mechanisms, the shorter the plasma half-life of the hormone.

Regulation of Hormone Levels

Hormone levels can be modulated through regulatory factors affecting any of the steps listed earlier; however, the bulk of the acute "fine-tuning" of hormone levels occurs at the level of hormone secretion and synthesis. Many, if not most, hormone levels are controlled either directly or indirectly by the biological activity that they serve to control. For example, parathyroid hormone (PTH) secretion, which responds to low extracellular calcium levels, mobilizes calcium out of bone which, in turn, signals back to the parathyroid gland to turn off additional PTH secretion. This negative feedback loop is a hallmark of endocrine regulation. The end product or negative regulator can either be an inorganic ion or metabolite (eg, calcium for PTH) or a hormonal product in the endocrine cascade (eg, thyroid hormone for TSH). Not all feedback is negative in nature. Positive feedback loops (eg, mid-cycle estradiol-induced luteinizing hormone secretion) also play important roles in governing physiological homeostasis.

HORMONE ACTION

Hormones produce their biologic effects through interaction with high-affinity receptors that are, in turn, linked to one or more effector systems within the cell. These effectors involve many different components of the cell's metabolic machinery, ranging from ion transport at the cell surface to stimulation of the nuclear transcriptional apparatus. Steroids and thyroid hormones exert their effects in the cell nucleus, although regulatory activity in the extranuclear compartment has also been documented. Peptide hormones and neurotransmitters, on the other hand, trigger a plethora of signaling activity in the cytoplasmic and membrane compartments while at the same time exerting parallel effects on the transcriptional apparatus. The discussion that follows will focus on the primary signaling systems employed by selected hormonal agonists and attempt to identify examples where aberrant signaling results in human disease.

RECEPTORS

The biologic activity of individual hormones is dependent on their interactions with specific high-affinity receptors on the surfaces or in the cytoplasm or nuclei of target cells. The receptors, in turn, are linked to signaling effector systems responsible for generating the observed biologic responses. Receptors, therefore, convey not only specificity of the response (ie, cells lacking receptors lack responsiveness to the hormone) but also the means for activating the effector mechanism. In general, receptors for the peptide hormones and neurotransmitters are aligned on the cell surface and those for the steroid hormones, thyroid hormone, and vitamin D are found in the cytoplasmic or nuclear compartments, although, as noted earlier, exceptions have been identified in both cases.

Interactions between the hormone ligand and its receptor are governed by the laws of mass action:

$$[H]+[R] \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} [HR]$$
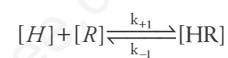
where $[H]$ is the hormone concentration, $[R]$ is the receptor concentration, $[HR]$ is the concentration of the hormone–receptor

complex, and k_{+1} and k_{-1} are the rate constants for [HR] formation and dissociation, respectively. Thus, at equilibrium,

$$k_{+1}[H][R] = k_{-1}[HR]$$

or

$$\frac{[H][R]}{[HR]} = \frac{k_{-1}}{k_{+1}} = K_D$$

where K_D is the equilibrium dissociation constant that defines the affinity of the hormone–receptor interaction (ie, lower the dissociation constant, higher the affinity). Assuming that total receptor concentration R_0 = [HR] + [R], this equation can be rearranged to give

$$\frac{[HR]}{[H]} = -\left(\frac{[HR]}{K_D}\right) + \frac{R_0}{K_D}$$

This is the Scatchard equation and states that when bound ligand over free ligand (ie, [HR]/[H]) is plotted against bound ligand (ie, [HR]), the slope of the line is defined by $-1/K_D$, the y-intercept by R_0/K_D, and the x-intercept by R_0 (Figure 1–3). When [HR] = $R_0/2$, [H] = K_D; therefore, the K_D is also the concentration of hormone [H]
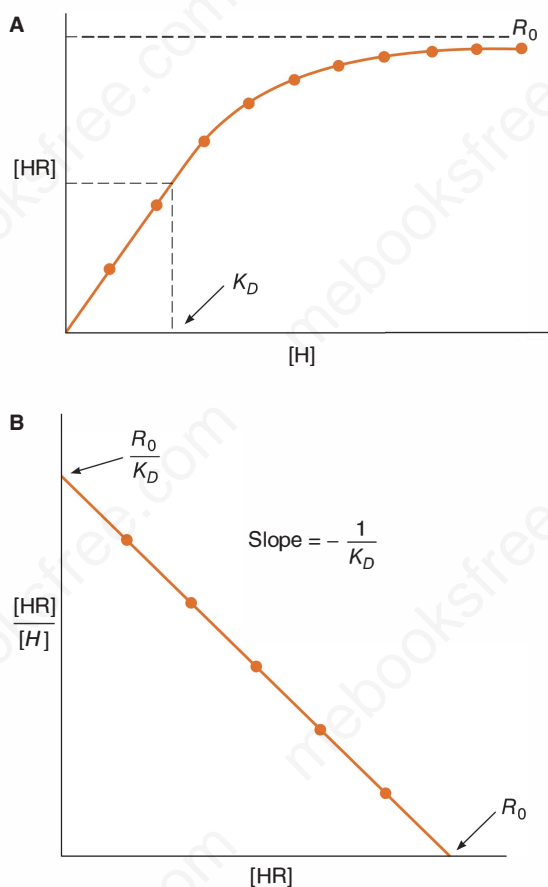
FIGURE 1–3 Ligand saturation (A) and Scatchard analysis (B) of a hypothetical hormone receptor interaction. K_D represents the dissociation constant; R_0 the total receptor concentration; [HR] and [H] the bound and free ligand, respectively. Note in (A) that the K_D is the concentration [H] at which half of available receptors are occupied.

at which one-half of the available receptors are occupied. Thus, knowledge of bound and free ligand concentrations, which can be determined experimentally, provides information regarding the affinity of the receptor for its ligand and the total concentration of receptor in the preparation.

Agents that bind to receptors with high affinity are classified as either agonists or antagonists based on the functional outcome of this receptor–ligand interaction. Agonists are ligands that trigger the effector mechanisms and produce biologic effects. Antagonists bind to the receptor but do not activate the effector mechanisms. Because they occupy receptor and block association with the agonist, they antagonize the functional activity of the agonist. Partial agonists bind to the receptor but possess limited ability to activate the effector mechanisms. In different circumstances, partial agonists may demonstrate variable biologic activity. For example, when used alone, they may display weak activating activity, whereas their use together with a full agonist may lead to inhibition of function because the latter is displaced from the receptor molecule by a ligand with lower intrinsic activity.

In some systems, receptors are available in surplus, which may be several-fold higher than that required to elicit a maximal biologic response. Although such spare receptor systems superficially appear redundant, they are designed to rectify a mismatch between low circulating ligand levels and a relatively low-affinity ligand–receptor interaction. Thus, by increasing the number of available receptors, the system is guaranteed a sufficient number of ligand-bound receptor units to activate downstream effector systems fully, despite operating at subsaturating levels of ligand.

NEUROTRANSMITTER AND PEPTIDE HORMONE RECEPTORS

As mentioned earlier, neurotransmitter and peptide hormones interact predominantly with receptors expressed on the plasma membrane at the cell surface. The K_D of a neurotransmitter for its receptor is typically higher than that of a hormone for its receptor, reflecting a higher k_{off} rate constant (see earlier). Neurotransmitter receptor occupancy is driven by the extraordinarily high concentrations of ligand that can be achieved in the synaptic cleft, and occupancy of the hormone receptor is driven by its high affinity for ligand. The high k_{off} of the neurotransmitter–receptor interaction guarantees that the effect is rapid in onset but of short duration, whereas the lower k_{off} of the hormone–receptor interaction guarantees that the effect is slow in onset but difficult to extinguish, kinetics that are more appropriate for the hormonal functions of these ligands.

The neurotransmitter and peptide receptors can be divided into several major groups (Table 1–1 and Figure 1–4). The first includes the so-called serpentine or "seven-transmembrane-domain" receptors. These receptors each contain an amino terminal extracellular domain followed by seven hydrophobic amino acid segments, each of which is believed to span the membrane bilayer (see Figure 1–4). The seventh of these, in turn, is followed by a hydrophilic carboxyl terminal domain that resides within the cytoplasmic compartment. As a group, they share a dependence

TABLE 1–1 Major subdivisions (with examples) of the neurotransmitter-peptide hormone receptor families.[a]

Seven-Transmembrane Domain
β-Adrenergic
PTH
LH
TSH
GRH
TRH
ACTH
MSH
Glucagon
Dopamine
α₂-Adrenergic (–)
Somatostatin (–)
Single-Transmembrane Domain
Growth factor receptors
Insulin
IGF
EGF
PDGF
Cytokine receptors
Growth hormone
Prolactin
Erythropoietin
CSF
Guanylyl cyclase-linked receptors
Natriuretic peptides

[a]Receptors have been subdivided based on shared structural and functional similarities. Minus (–) sign denotes a negative effect on cyclase activity.

on the G protein transducers (GPCRs discussed later) to execute many of their biologic effects. A second group includes the single-transmembrane-domain receptors that harbor intrinsic tyrosine kinase activity. This includes the insulin, insulin-like growth factor (IGF), and epidermal growth factor (EGF) receptors. A third group, which is functionally similar to the second group, is characterized by a large, extracellular binding domain followed by a single membrane-spanning segment and a cytoplasmic tail. These receptors do not possess intrinsic tyrosine kinase activity but appear to function through interactions with soluble transducer molecules which do possess such activity. Prolactin and growth hormone are included in this group. A fourth group is the transforming growth factor beta (TGF-β) family which signals through serine/threonine kinase domains in their cytoplasmic tails. A fifth group, which includes the natriuretic peptide receptors, operates through activation of a particulate guanylyl cyclase and synthesis of cGMP. The cyclase is covalently attached at the carboxyl terminal portion of the ligand-binding domain (LBD) and thus represents an intrinsic part of the receptor molecule.

G PROTEIN–COUPLED RECEPTORS

G protein–coupled receptors (GPCRs) constitute a large superfamily of molecules capable of responding to ligands of remarkable structural diversity—ranging from photons to large polypeptide hormones. Because of their diversity, GPCRs are the target of over 40% of modern pharmaceuticals. GPCRs initiate
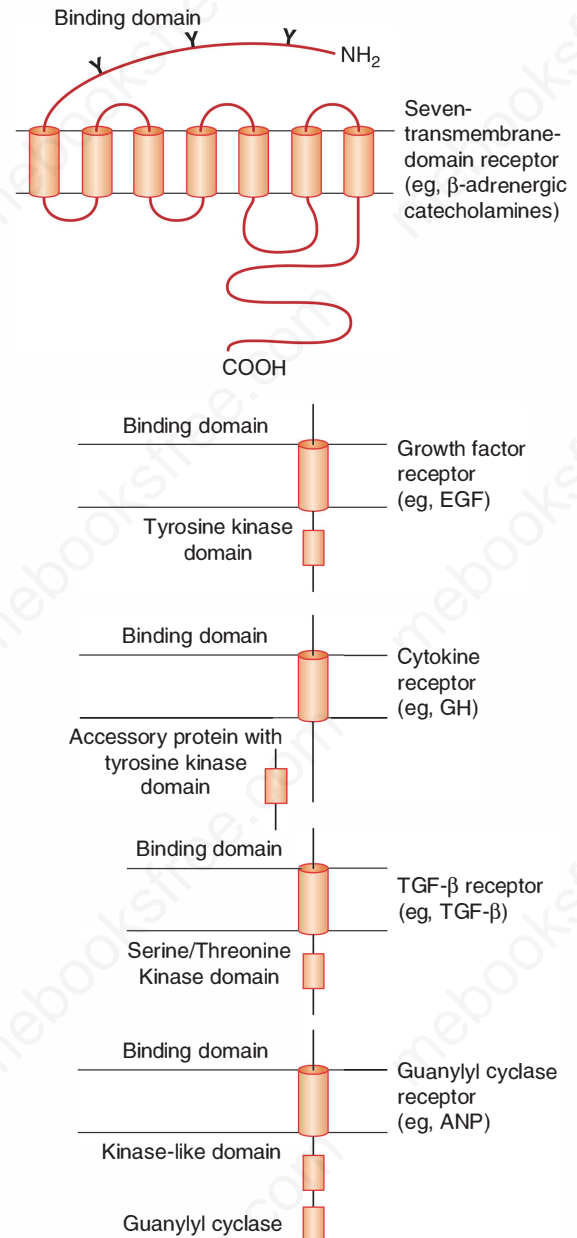
FIGURE 1–4 Structural schematics of different classes of membrane-associated hormone receptors. Representative ligands are presented in parentheses (ANP, atrial natriuretic peptide; EGF, epidermal growth factor; GH, growth hormone; TGF-β, transforming growth factor beta).

intracellular signaling by activating one (or in some cases multiple) G proteins resulting in biological responses. These receptors share overall structural features, most notably seven membrane-spanning regions connected by intracellular and extracellular loops (see Figure 1–4). The receptors are oriented such that the amino terminal domain is extracellular, whereas the carboxyl terminal tail is cytoplasmic. The membrane-spanning segments interact with one another, forming an irregular cylindrical bundle around a central cavity within the molecule. GPCRs can assume at least two conformations with differing orientations of the

membrane-spanning segments relative to one another. One orientation is favored in the absence of an agonist ligand. In this orientation the receptor does not activate a G protein (inactive conformation). The second orientation is stabilized by the binding of an appropriate agonist ligand. In this conformation the receptor activates a cognate G protein (active conformation). All GPCRs are thought to undergo a similar conformational switch on agonist binding, producing a structural change in the cytoplasmic domain that promotes G protein activation. Some small agonists, such as catecholamines, are able to enter the cavity formed by the transmembrane segments, thereby directly stabilizing the active receptor conformation. Other agonists, such as large polypeptide hormones, bind primarily to the extracellular domain of their GPCRs. More recently, a number of orphan GPCRs have been found to be activated by hydrophobic ligands including steroids (eg, estrogen binding to GPR30) and lipids (eg, LPA binding to GPR23). Ligand binding indirectly results in movement of the transmembrane region of the receptor and stabilization of the active receptor conformation.

Until recently, it was thought that GPCRs function exclusively as monomers. Many GPCRs are now known to dimerize either with themselves (homodimerization) or with other GPCRs (heterodimerization). In some cases, dimerization is important for efficient receptor biosynthesis and membrane localization. In other cases, dimerization is important for optimal ligand affinity, specificity, or receptor signaling.

Heritable mutations in a variety of GPCRs are known to be associated with disease. Loss-of-function phenotypes can result from mutations that eliminate one or both receptor alleles, or that result in the synthesis of signaling-defective receptors. Gain-of-function phenotypes generally result from point mutations that produce constitutively active receptors (ie, stably assume the active receptor conformation even in the absence of an agonist ligand). Examples of such GPCR disorders relevant to endocrinology are described later and discussed in greater detail elsewhere in this book.

G PROTEIN TRANSDUCERS

G proteins are a family of heterotrimeric proteins that regulate the activity of effector molecules (eg, enzymes, ion channels) (examples in Table 1–2), ultimately resulting in biological responses. The identity of a G protein is defined by the nature of its α subunit, which is largely responsible for effector activation. The major G proteins involved in hormone action (and their actions on effectors) are G_s (stimulation of adenylyl cyclase), G_i (inhibition of adenylyl cyclase; regulation of calcium and potassium channels), and $G_{q/11}$ (stimulation of phospholipase C [PLC] β). Recently, GPCRs linked to $G_{12/13}$ were identified as key inputs of the Hippo/YAP/TAZ transcriptional regulators, which play a central role in controlling organ size, growth, and integrating extracellular cues. In each of these cases, the β and γ subunits of G proteins are tightly associated with one another and function as a dimer. In some cases, the $\beta\gamma$ subunit dimer can also regulate effector function.

TABLE 1–2 G protein subunits selectively interact with specific receptor and effector mechanisms.

G Protein Subunit	Representative Associated Receptors	Effector
α_s	β-Adrenergic TSH Glucagon	Adenylyl cyclase Ca^{2+} channels K^+ channels
α_i	α_2-Adrenergic Muscarinic (type II)	Adenylyl cyclase Ca^{2+} channels K^+ channels
α_q	α_1-Adrenergic	PLCβ
β/γ		Adenylyl cyclase (+ or −) PLC Supports βARK-mediated receptor phosphorylation and desensitization

G proteins are noncovalently tethered to the plasma membrane and are thus proximate to their cognate receptors and to their effector targets. The basis for specificity in receptor–G protein interactions has not been fully defined. It is likely that specific structural determinants presented by the cytoplasmic loops of the GPCR determine the identity of the G proteins that are activated. It is the nature of the α subunit of the G protein that is critical for receptor signaling. There are about a dozen different G protein α subunits and hundreds of distinct GPCRs.

Clearly, each specific G protein can be activated by a large number of different receptors. For example, G_s is activated by receptors for ligands as diverse as β-adrenergic catecholamines and large polypeptide hormones such as luteinizing hormone (LH). LH is thereby able to stimulate adenylyl cyclase and raise intracellular levels of cAMP in cells that express LH receptors (eg, Leydig cells of the testis). In contrast, an individual GPCR can couple to multiple Gα subunits, often in response to different ligands (eg, PTH receptor can activate G_s, G_i, and G_q).

Figure 1–5 is a schematic representation of the molecular events associated with activation of G proteins by GPCRs. In the basal, inactive state, the G protein is an intact heterotrimer with guanosine diphosphate (GDP) bound to the α subunit. Agonist binding to a GPCR promotes the physical interaction between the receptor and its cognate G protein. This produces a conformational change in the G protein, resulting in the dissociation of GDP. This in turn allows the binding of GTP (which is present at a much higher concentration in cells than is GDP) to the α subunit. Dissociation of the GTP-bound α subunit from the $\beta\gamma$ dimer then occurs, allowing these subunits to activate their effector targets. Dissociation of the hormone–receptor complex also occurs. The duration of activation is determined by the intrinsic GTPase activity of the G protein α subunit. Hydrolysis of GTP to GDP terminates the activity and promotes reassociation of the $\alpha\beta\gamma$ trimer, returning the system to the basal state. The GTPase activity of G protein α subunits can be increased by the action of proteins termed "regulators of G protein signaling" (RGS proteins) which act by increasing the speed of GTP cycling.
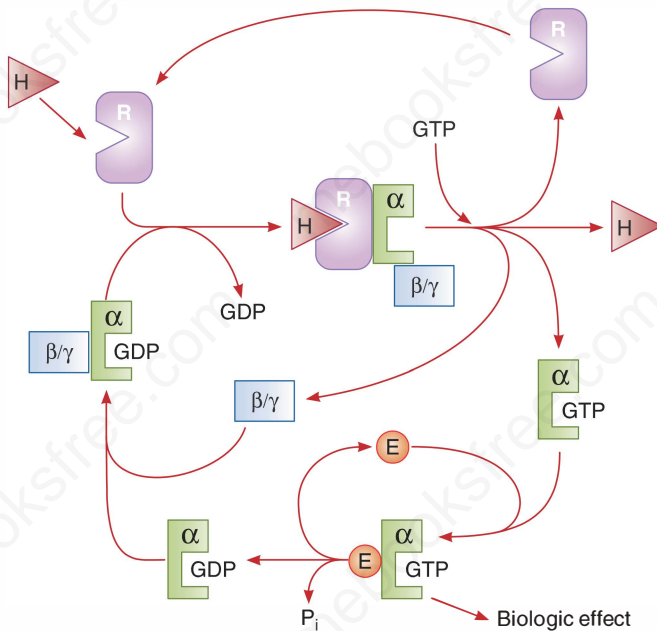
FIGURE 1–5 G protein–mediated signal transduction. α and β/γ subunits of a representative G protein are depicted (see text for details) (E, effector; H, hormonal ligand; R, hormone receptor).

EFFECTORS

Numerous effectors have been linked to the GPCRs. A number of these are presented in Table 1–2. A great many other G proteins—not dealt with here—are coupled to physical or biochemical stimuli but have very limited involvement in hormone action. As discussed, adenylyl cyclase, perhaps the best studied of the group, is activated by G_s (Figure 1–6). This activation results in a transient increase in intracellular cAMP levels. The cAMP binds to the inhibitory regulatory subunit of inactive protein kinase A (PKA) and promotes its dissociation from the complex, thereby permitting enhanced activity of the catalytic subunit. The latter phosphorylates a variety of cellular substrates, among them the hepatic phosphorylase kinase that initiates the enzymatic cascade which results in enhanced glycogenolysis. It also phosphorylates and activates the cAMP response element–binding protein (CREB), which mediates many of the known transcriptional responses to cAMP (and to some extent calcium) in the nuclear compartment. Other transcription factors are also known to be phosphorylated by PKA.

PLC beta (PLCβ) is a second effector system that has been studied extensively. The enzyme is activated through G_q-mediated
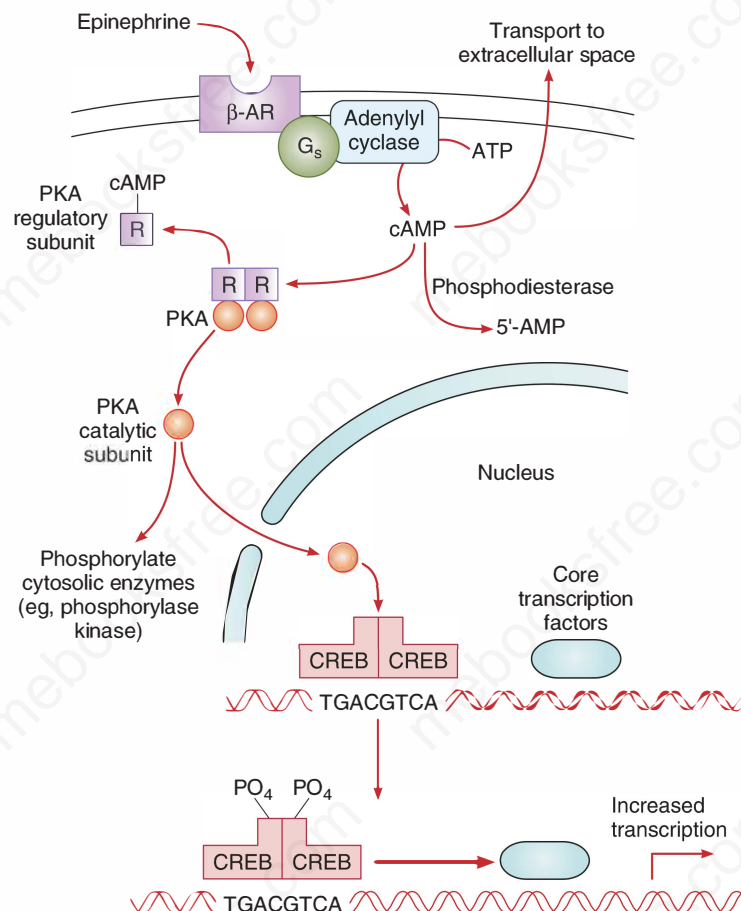
FIGURE 1–6 β-Adrenergic receptor/G_s mediated signaling in the cytoplasmic and nuclear compartments. The cAMP response element–binding protein (CREB) is depicted bound to a consensus CRE in the basal state. Phosphorylation of this protein leads to activation of the juxtaposed core transcriptional machinery.

transduction of signals generated by a wide array of hormone–receptor complexes, including those for angiotensin II, α-adrenergic agonists, and endothelin. Activation of the enzyme leads to cleavage of phosphoinositol 4,5-bisphosphate in the plasma membrane to generate inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG) (Figure 1–7). The former interacts with a specific receptor present on the endoplasmic reticulum membrane to promote release of Ca^{2+} into the cytoplasmic compartment. The increased calcium, in turn, may activate protein kinases, promote secretion, or foster contractile activity. Depletion of intracellular calcium pools by IP_3 results in enhanced uptake of calcium across the plasma membrane (perhaps through generation of IP_4 [1,3,4,5-tetrakisphosphate]), thereby activating a second, albeit indirect, signaling mechanism that serves to increase intracellular calcium levels even further. DAG functions as an activator of several protein kinase C (PKC) isoforms within cells. Several different isoforms of PKC (eg, α, β, γ) may exist in a given cell type. A number of these are calcium-dependent, a property which, given the IP_3 activity mentioned earlier, provides the opportunity for a synergistic interaction of the two signaling pathways driven by PLCβ activity.

However, not all PKC activity derives from the breakdown of PIP_2 substrate. Metabolism of phosphatidylcholine by PLC_{PC} (phosphatidylcholine-selective phospholipase) leads to the generation of phosphocholine and DAG. This latter pathway is believed to be responsible for the more protracted elevations in PKC activity seen following exposure to agonist.

Other phospholipases may also be important in hormone-dependent signaling. Phospholipase D employs phosphatidylcholine as a substrate to generate choline and phosphatidic acid. The latter may serve as a precursor for subsequent DAG formation. As with PLC_{PC} earlier, no IP_3 is generated as a consequence of this reaction. Phospholipase A_2 triggers release of arachidonic acid, a precursor of prostaglandins, leukotrienes, endoperoxides, and thromboxanes, all signaling molecules in their own right. The relative contribution of these other phospholipases to hormone-mediated signal transduction and the role of the specific lipid breakdown products (eg, phosphocholine, phosphatidic acid) in conveying regulatory information remains an area of active research.

Activation of effectors by GPCRs is subject to regulatory mechanisms that prevent overstimulation of cells by an agonist
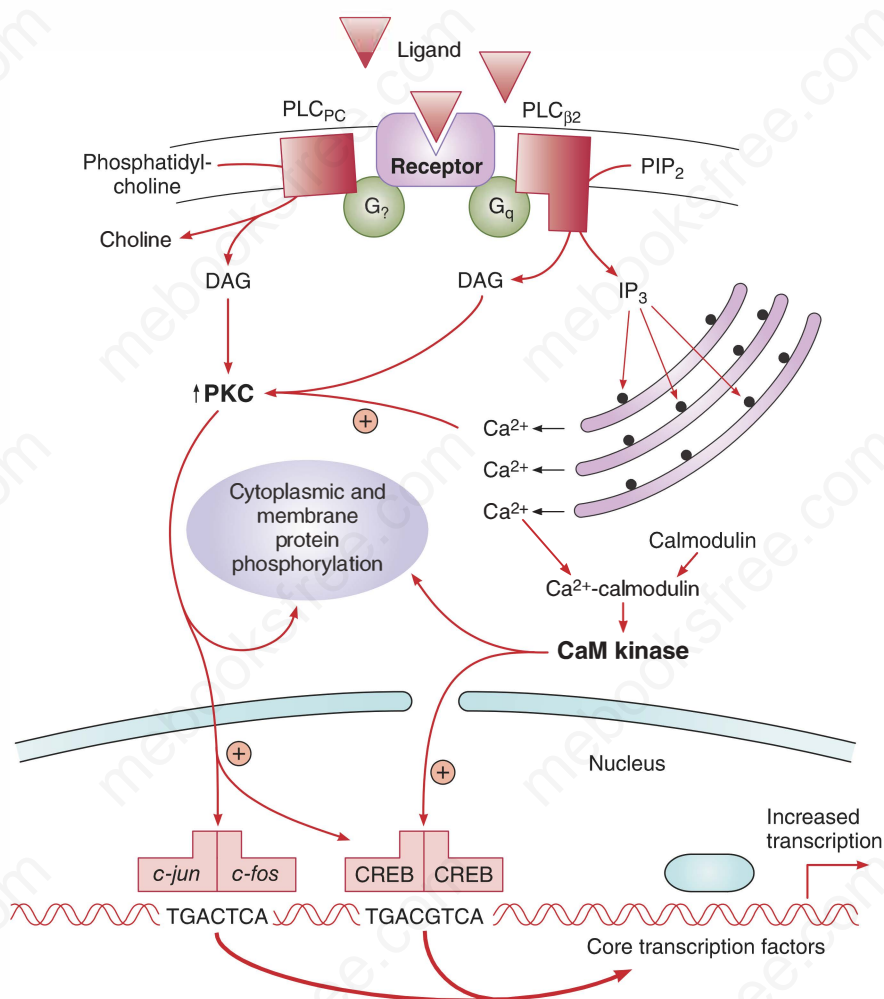
FIGURE 1–7 PLCβ-coupled receptor/G_q mediated signaling in the cytoplasmic and nuclear compartments (DAG, diacylglycerol; PC, phosphatidylcholine; PKC, protein kinase C; PLC, phospholipase).
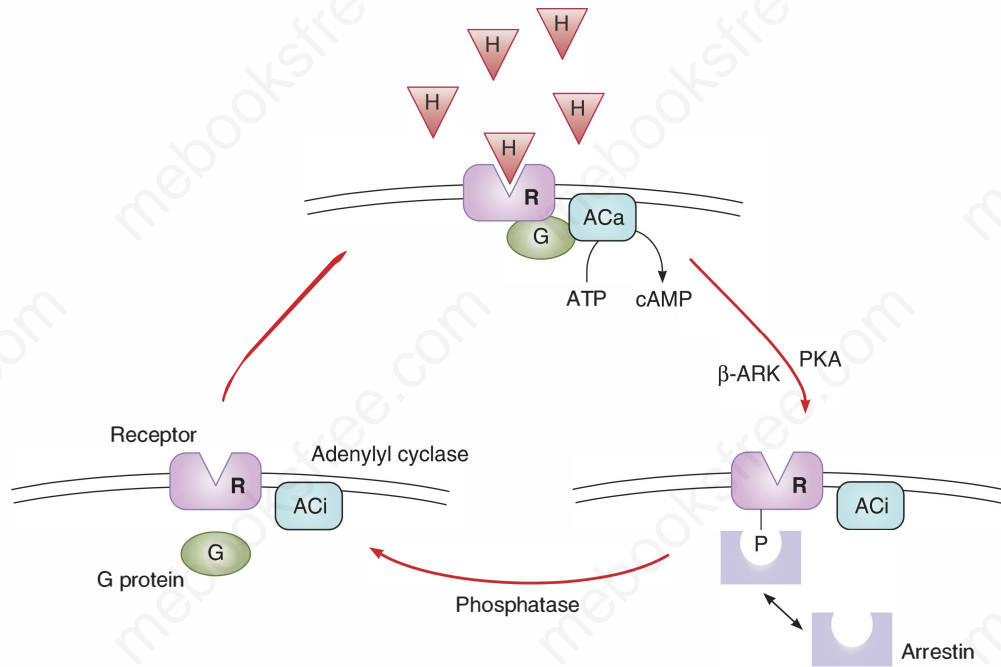
FIGURE 1–8 Kinase-dependent desensitization of the ligand–receptor complex. Schema shown is that for the β-adrenergic receptor, but similar systems probably exist for other types of G protein–linked receptors (ACa, active adenylyl cyclase; ACi, inactive adenylyl cyclase; β-ARK, β-adrenergic receptor kinase; PKA, protein kinase A).

ligand. At the level of the receptor, two regulatory events are known to occur. One is desensitization, wherein initial stimulation of a receptor by its agonists leads to a loss of the ability of the receptor to subsequently elicit G protein activation. This is shown schematically in Figure 1–8 for the β-adrenergic receptor. A similar regulatory mechanism exists for many other GPCRs. Agonist binding to the receptor produces G protein activation and results also in activation of a kinase (termed G protein–coupled receptor kinase, GRK) that phosphorylates the cytoplasmic domain of the receptor. By virtue of this phosphorylation, the receptor acquires high affinity for a member of the arrestin family of proteins. The name "arrestin" derives from the observation that the receptor is no longer capable of interacting with a G protein when arrestin is bound. Thus, the phosphorylated receptor becomes uncoupled from its G protein, preventing signaling to the effector. The receptor remains inactive until a phosphatase acts to restore the receptor to its unphosphorylated state thus releasing the bound arrestin. Recently, β arrestin-dependent signaling pathways have been identified where the arrestins can link GPCRs to intracellular signaling such as the MAPK cascades. Beta arrestins also appear to act directly as signal transducers by interacting with regulators of transcription factors. This receptor-independent signaling capacity likely has implications for endocrine diseases including those affecting bone.

Many GPCRs are also susceptible to agonist-induced down-regulation, resulting in a reduced level of cell surface receptors following exposure of cells to an agonist. This can result from agonist-induced internalization of receptors, followed by trafficking of receptors to lysosomes where degradation occurs. In addition, chronic exposure of cells to an agonist may result in signaling events that suppress the biosynthesis of new receptors, thereby lowering steady-state receptor levels. Together, these regulatory events ensure that the cell is protected from excessive stimulation in the presence of sustained high levels of an agonist.

Recently, it has become clear that these events serving to dampen G protein signaling can also have important roles in promoting cell signaling. For example, arrestin association with GPCRs can activate specific pathways such as the MAP kinase pathway independently of G protein signaling. In addition, internalized GPCRs can, in some cases, retain the ability to signal, and the effects may differ from those produced when activation occurs at the plasma membrane.

DISORDERS OF G PROTEINS AND G PROTEIN–COUPLED RECEPTORS

Two bacterial toxins are capable of covalently modifying specific G protein α subunits, thereby altering their functional activity. Cholera toxin is a protein that binds to receptors present on all cells, resulting in the internalization of the enzymatic subunit of the toxin. The toxin enzyme is an ADP-ribosyl transferase that transfers ADP-ribose from NAD to an acceptor site (Arg^{201}) on the α subunit of G_s. This covalent modification greatly inhibits the GTPase activity of α_s, enhancing the activation of adenylyl cyclase by extending the duration of the active GTP-bound form of the G protein. Even in the absence of an active GPCR, GDP dissociates (albeit very slowly) from the G protein. Thus, cholera toxin will eventually activate adenylyl cyclase activity even without agonist binding to a GPCR. The result is a large and sustained activation of adenylyl cyclase. When this occurs in intestinal epithelial cells,

the massive increase in cAMP results in the increased water and salt secretion into the gut that is characteristic of cholera.

Pertussis toxin is also an ADP-ribosyl transferase. However, in this case, the substrates are α subunits of different G proteins, most notably G_i and G_o. The ADP-ribose moiety is transferred to a cysteine residue near the carboxyl terminus of the α subunit, a region required for interaction with activated GPCRs. Once ADP-ribosylated by pertussis toxin, these G proteins are no longer able to interact with activated receptors and are thus stuck in an inactive (GDP-bound) conformation. Inhibition of receptor-mediated activation of G_i and G_o accounts for many of the clinical manifestations of pertussis infection.

Genetic mutations in G protein α subunits are seen in a number of human diseases. Acquired activating mutations in $G\alpha_s$ can produce a variety of phenotypes depending on the site of expression of the mutant protein. In McCune-Albright syndrome, the mutation occurs in a subset of neural crest cells during embryogenesis. All of the descendants of these cells, including certain osteoblasts, melanocytes, and ovarian or testicular cells, express the mutant protein. The result is a form of genetic mosaicism in which the consequence of unregulated production of cAMP in particular tissues is evident (ie, the progressive bone disorder polyostotic fibrous dysplasia, the abnormal skin pigmentation referred to as café au lait spots, and gonadotropin-independent precocious puberty). In cells where cAMP is linked to cell proliferation (eg, thyrotropes, somatotropes), a subset of patients with benign tumors has been shown to have acquired activating mutations in $G\alpha_s$. Activating mutations in $G\alpha_{i2}$, which is coupled to cell proliferation, have been reported in a subset of adrenal and ovarian tumors.

Loss-of-function mutations in $G\alpha_s$ are associated with the hereditary disorder pseudohypoparathyroidism type 1 (PHP-1a). This disorder, first described by Fuller Albright, was the first documented example of a human disease attributable to target cell resistance to a hormone. Affected patients display biochemical features of hypoparathyroidism (eg, hypocalcemia, hyperphosphatemia) but have markedly increased circulating levels of PTH and display target cell resistance to PTH. Many hormone receptors couple to adenylyl cyclase via $G\alpha_s$, yet patients with PHP-1a generally display only subtle defects in responsiveness to other hormones (eg, TSH, LH). The explanation for this lies in the fascinating genetics of this disorder. In brief, affected patients have one normal and one mutated $G\alpha_s$ allele. The mutated allele fails to produce an active form of the protein. Tissues in these patients are expected to express about 50% of the normal level of $G\alpha_s$, a level sufficient to support signaling to adenylyl cyclase. However, in certain tissues, the α_s gene is subject to genetic imprinting such that the paternal allele is expressed poorly or not at all. In individuals harboring inactivating mutations, if the paternal allele has the mutation, all cells express about 50% of the normal level of $G\alpha_s$ (derived from the normal maternal allele). However, if the maternal allele has the mutation, then the cells in which paternal imprinting occurs express low levels or no $G\alpha_s$. One of the major sites of this paternal imprinting is in the proximal renal tubule, an important target tissue for the physiologic actions of PTH. This accounts for the clinical resistance to PTH seen in PHP-1a and accounts also for the fact that only a subset of patients

with haploinsufficiency of α_s are resistant to PTH. Interestingly, essentially all patients with haploinsufficiency of $G\alpha_s$ display Albright hereditary osteodystrophy, a developmental disorder with phenotypic manifestations affecting a variety of tissues. This indicates that even a partial loss of adenylyl cyclase signaling is incompatible with normal development.

Mutations in the genes encoding GPCRs are being increasingly recognized as important in the pathogenesis of endocrine disorders. Loss-of-function mutations generally need to be homozygous (or compound heterozygous) in order to result in a significant disease phenotype. This is probably due to the fact that most cells express higher levels of each receptor, above what is needed for maximal cellular response (spare receptors). Thus, a 50% reduction in the amount of a cell surface receptor may have little influence on the ability of a target cell to respond. However, in some situations, haploinsufficiency of a GPCR can produce a clinical phenotype. For instance, heterozygous loss-of-function mutations in the G protein–coupled calcium-sensing receptor results in the autosomal dominant disorder familial hypocalciuric hypercalcemia type 1 due to usually mild dysregulation of PTH secretion and renal calcium handling. Homozygous loss-of-function of the calcium-sensing receptor results in severe neonatal hyperparathyroidism due to the loss of the ability of plasma calcium to suppress PTH secretion and promote renal calcium clearance. Syndromes of hormone resistance have also been reported in patients lacking expression of functional GPCRs for vasopressin, ACTH, and TSH. Loss of functional expression of the PTH receptor results in Blomstrand chondrodysplasia, a disorder that is lethal due to the inability of PTH-related protein (a PTH receptor agonist) to promote normal cartilage development.

Mutations that render GPCRs constitutively active (in the absence of an agonist ligand) are seen in a number of endocrine disorders. Generally speaking, such mutations produce a disease phenotype resembling that seen with excessive levels of the corresponding hormone agonist. Thus, activating mutations in the TSH receptor produce neonatal thyrotoxicosis, and activating mutations in the LH receptor result in pseudoprecocious puberty or testotoxicosis. Activating mutations in the PTH receptor result in Jansen-type metaphyseal chondrodysplasia; a disorder characterized by hypercalcemia and increased bone resorption (mimicking the effects of excess PTH on bone) and delayed cartilage differentiation (mimicking the effects of excess PTH-related protein on cartilage). An approach to treating disorders resulting from constitutively active GPCRs would be administration of "inverse agonists," agents that stabilize receptors in their inactive conformation. Although inverse agonists have been identified for a number of GPCRs, they have yet to be successfully employed as therapeutics. In contrast, molecular mimics of endogenous ligands have found utility as a way to stimulate signaling *via* allosteric receptor changes. This is the basis for cinacalcet's activity as a calcimimetic on the calcium-sensing receptor in the parathyroid cell and accounts for its utility in treating secondary hyperparathyroidism. Finally, molecular analysis of GPCRs has revealed that point mutations, in addition to producing constitutive activity, can alter the specificity of ligand binding or the ability of the receptor to become desensitized. It is almost certain that such mutations will be found to provide the basis for some more subtle endocrinopathies.

GROWTH FACTOR RECEPTORS

The growth factor receptors differ from those described earlier both structurally and functionally. Unlike the GPCRs, the growth factor receptors span the membrane only once and acquire their signaling ability, at least in part, through activation of tyrosine kinase activity, which is intrinsic to the individual receptor molecules. The insulin and IGF receptors fall within this group, as do those for the autocrine or paracrine regulators platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), and EGF. Signaling is initiated by the association of ligand (eg, insulin) with the receptor's extracellular domain (Figure 1–9) and subsequent
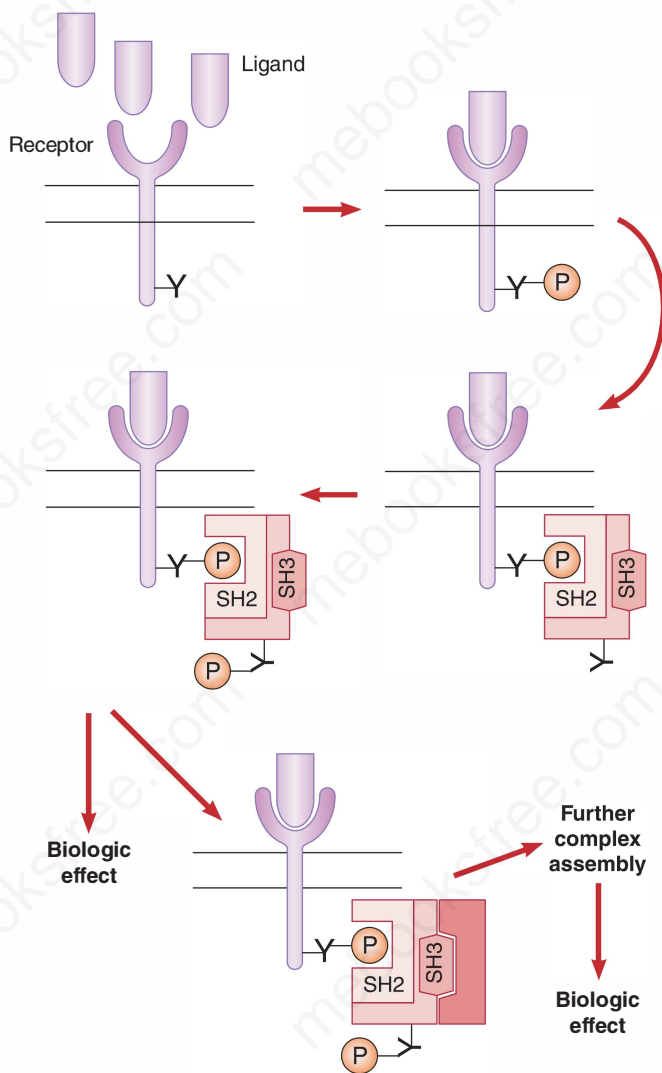
FIGURE 1–9 Signaling by a tyrosine kinase–containing growth factor receptor. Receptors depicted here as monomers for simplicity; typically dimerization of receptors follows association with ligand. Autophosphorylation of one or more critically positioned tyrosine residues in the receptor leads to association with accessory proteins or effectors through SH2 domains present on the latter. In some cases, an SH3 domain present on the same protein leads to recruitment of yet other proteins leading to further complex assembly.

receptor dimerization. The duration of signaling can be regulated by clathrin, a protein that is required for cellular endocytosis. Internalization of the ligand/receptor complex results in phosphorylation of tyrosines both on the receptor itself as well as on nonreceptor substrates. It is assumed that phosphorylation of these substrates results in a cascade of activation events, similar to those described for the G protein–coupled systems, which contribute to perturbations in intracellular pathways. The autophosphorylation of the receptor molecules themselves has been studied extensively and provide some intriguing insights into the mechanisms that underlie signal transduction by this group of proteins.

Tyrosine phosphorylation takes place at specific locations in the receptor molecule. Once phosphorylated, these sites associate, in highly specific fashion, with a variety of accessory proteins that possess independent signaling capability. These include PLCγ, phosphoinositol (PI) 3′ kinase, GTPase-activating protein (GAP), growth factor receptor–bound protein-2 (GRB2), and the Src family nonreceptor tyrosine kinases. These interactions are fostered by the presence of highly conserved type 2 *src* homology (based on sequence homology to the *src* proto-oncogene) domains (SH2) in each of the accessory molecules. Each individual SH2 domain displays specificity for the contextual amino acids surrounding the phosphotyrosine residues in the receptor molecule. In the PDGF receptor, for example, the SH2 domain of PLCγ associates selectively with Tyr^{977} and Tyr^{989}, whereas that of PI 3′ kinase associates with Tyr^{708} and Tyr^{719}. Thus, diversity of response is controlled by contextual sequences around individual phosphotyrosine residues that determine the types of accessory proteins brought into the signaling complex. These protein–protein interactions may provide a means of directly controlling the signaling molecule in question, perhaps through a change in steric conformation. Alternatively, they may facilitate the sequestration of these accessory proteins in or near the plasma membrane compartment, in close proximity to key substrates (eg, membrane lipids in the case of PLCγ) or other important regulatory proteins.

Some of these associations trigger immediate signaling events, but others (eg, GRB2) may act largely to provide the scaffolding needed to construct a more complex signaling apparatus (Figure 1–10). In the case of GRB2, another accessory protein (son-of-sevenless; SOS) associates with the receptor–GRB2 complex through a type 3 *src* homology (SH3) domain present in the latter. This domain recognizes a sequence of proline-rich amino acids present in the SOS protein. SOS, in turn, facilitates assembly of the Ras–Raf complex, which permits activation of downstream effectors such as mitogen-activated protein kinase (MAPK) kinase (MEK). This latter kinase, which possesses both serine-threonine and tyrosine kinase activity, activates the p42 and p44 MAPKs (also called extracellular signal–regulated kinases; ERKs). ERK acts on a variety of substrates within the cell, including the RSK kinases, which, in turn, phosphorylate the ribosomal S_6 protein and thereby stimulate protein synthesis. These phosphorylation reactions (and their amplification in those instances where the MAPK substrate is a kinase itself) often lead to protean changes in the phenotype of the target cells.
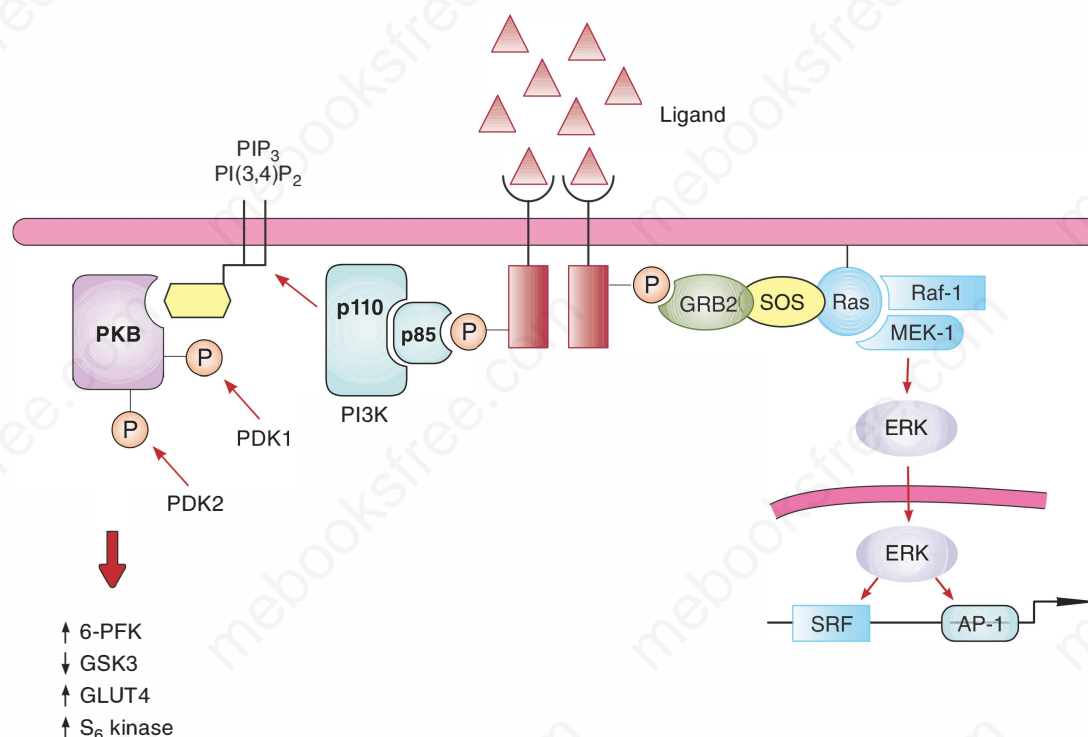
FIGURE 1–10 Growth factor–dependent pathway. Assembly of the components involved in the Ras/Raf/MEK/MAPK and PI-3K/PKB– signaling mechanisms.

The ligand bound growth factor receptors, including the insulin receptor, may also signal through the phosphoinositide 3-OH kinase (PI-3K). SH2 domains of the p85 regulatory subunit of PI-3K associate with the growth factor receptor through specific phosphotyrosine residues (Tyr^{740} and Tyr^{751} in the PDGF receptor) in a manner similar to that described earlier for GRB2 (see Figure 1–10). This leads to activation of the p110 catalytic subunit of PI-3K and increased production of phosphatidylinositol-3,4,5-trisphosphate (PIP_3) and phosphatidylinositol-3,4-bisphosphate ($PI[3,4]P_2$). These latter molecules sequester protein kinase B (also known as Akt) at the cell membrane through association with the plekstrin homology domains in the amino terminal of the kinase molecule. This in turn leads to phosphorylation of PKB at two separate sites (Thr^{308} in the active kinase domain and Ser^{473} in the carboxyl terminal tail) by PIP_3-dependent kinases (PDK1 and PDK2). These phosphorylations result in activation of PKB. In the case of insulin-sensitive target cells, downstream targets of activated PKB (eg, following insulin stimulation) include 6-phosphofructo-2-kinase (increased activity), glycogen synthase kinase-3 (GSK3) (decreased activity), the insulin-responsive glucose transporter GLUT 4 (translocation and increased activity), and p70 S6 kinase (increased activity). This leads to increased glycolysis, increased glycogen synthesis, increased glucose transport, and increased protein synthesis, respectively. This leads to increased glycolysis, increased glycogen synthesis, increased glucose transport, and increased protein synthesis, respectively. There is also a growing body of evidence suggesting that PKB may protect cells from programmed cell death through phosphorylation of key proteins in the apoptotic pathway.

It has been reported that GPCRs may also activate the Raf-MEK-ERK cascade, although in this case the signal traffics through a nonreceptor protein tyrosine kinase (NRPTK such as Src and Fyn) rather than the traditional growth factor receptor–linked tyrosine kinases. The details of the mechanism are incompletely understood, but it appears to require the participation of β-arrestin (discussed earlier) as an adaptor molecule linking the G protein receptor to the NRPTK. Interestingly, this implies that β-arrestin, which normally terminates coupling between the receptor and G protein, can actually promote coupling between the desensitized receptor and downstream effectors traditionally associated with growth factor–dependent activation.

CYTOKINE RECEPTORS

These include the receptors for a variety of cytokines, such as erythropoietin, colony-stimulating factor, GH, and prolactin. These cell membrane receptors have a single internal hydrophobic stretch of amino acids, suggesting that they span the membrane once (see Figure 1–4). They can be composed of monomers or heterodimers of different molecules.

Growth Hormone and Prolactin Receptors

Receptors for GH and prolactin are prototypical cytokine receptors (Figure 1–11). Interestingly, alternative splicing of the GH receptor gene primary transcript results in a foreshortened "receptor" that lacks the membrane anchor and carboxyl terminal domain of the protein. This "receptor" is secreted and serves to bind GH in the extracellular space (eg, circulating plasma). Unlike
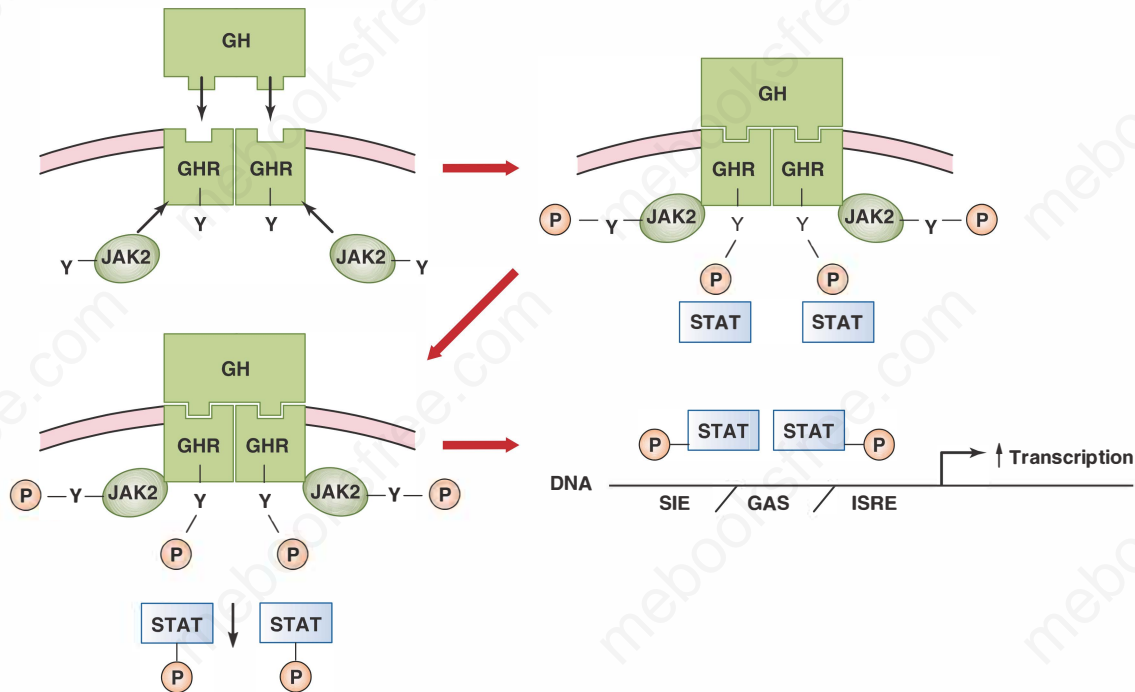
FIGURE 1–11 Signaling by the growth hormone receptor (GHR). Different portions of a single growth hormone molecule associate with homologous regions of two independent GHR molecules. This is believed to lead to the recruitment of JAK2, which phosphorylates the GHR, providing a docking site for STAT. The latter is phosphorylated, dissociates from the liganded receptor complex, and migrates to the nucleus, where it associates with binding elements of target genes and regulates transcription.

the growth factor receptors described earlier, GH receptors lack a tyrosine kinase domain. Different domains of a single GH molecule associate with homologous regions of two independent GH receptors, promoting dimerization of the receptors and subsequent association with and activation of Janus kinase (JAK) 2. Janus kinase 2 (JAK2) undergoes autophosphorylation and concurrently tyrosine phosphorylates the GH receptors. The latter provides a docking site for the signal transducer and activator of transcription (STAT) factors; STAT 5a and 5b appear to be particularly relevant to GH and prolactin action. The STATs are phosphorylated, dissociate from the GH receptor, migrate to the nucleus, and bind to specific STAT-binding DNA regulatory elements (SIE/ISRE/GAS) responsible for transcriptional control of GH target genes such as IGF-1. There are a number of different STAT family members, and there is specificity of certain cytokine receptors for certain STAT family members. This helps direct the specificity of signaling by each type of receptor. STAT signaling is also regulated by a family of inhibitors referred to as suppressor of cytokine signaling (SOCS) proteins. SOCS proteins bind to JAK and STAT proteins and target them for degradation. SOCS proteins are induced after cytokine/hormone binding and help to autoregulate signaling in this pathway.

TGF-β Receptors

These receptors bind a variety of ligands that include the cytokine transforming growth factor beta (TGF-β), the hormones inhibin, activin, anti-müllerian hormone (AMH), and the

bone morphogenetic protein (BMP) family. Diseases associated with mutations in the TGF-β receptor pathways can be very dramatic. For example, activating mutations in the type 1 activin A receptor ACVR1 result in severe heterotopic ossification in a disorder known as fibrodysplasia ossificans progressiva (FOP). The ligands for these receptors are typically homo- or heterodimers of subunits that have a highly conserved, cysteine-dependent structure. TGF-β family receptors bind to ligands through a heterodimeric receptor consisting of two transmembrane subunits known as type I and type II receptors (Figure 1–12). There are a number of different type I and type II receptor subunits in this family and type I/type II pairs can form between several different family members. Both type I and type II receptors have an intracellular serine/threonine kinase domain. The type II receptor is constitutively phosphorylated and active, while the type I receptor is not. The ligands in this family initially bind the type II receptor. The type I receptor is then recruited to the complex where the type II receptor kinase phosphorylates and activates the type I receptor, which then further propagates the signal. Downstream in the signaling pathway are a group of phosphorylation targets called the Smad proteins. These proteins, upon phosphorylation, can migrate to the nucleus to activate and/or repress transcription of target genes.

Modulators of TGF-β signaling also play critical roles in human disease. Members of the differential screening-selected gene in neuroblastoma (DAN) family were originally identified as BMP inhibitors and comprise a diverse group of antagonists.
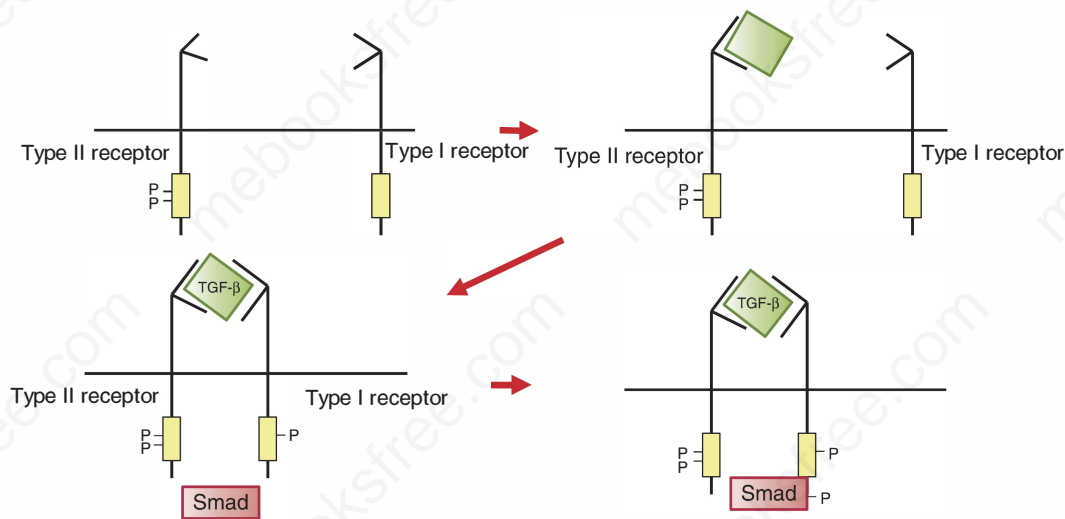
FIGURE 1–12 Signaling by the TGF-β receptors. TGF-β ligand first docks with the type II receptor which has an intracellular serine/threonine kinase domain that is constitutively active. After ligand binding, the type I receptor is then recruited to the complex and the type II receptor can phosphorylate the intracellular serine/threonine kinase domain of the type I receptor. This then propagates the signal downstream leading to the phosphorylation and activation of Smad proteins which can migrate to the nucleus and activate or repress gene transcription.

One DAN protein, sclerostin (SOST), is of particular interest because of its key role in regulating osteoblast anabolic function through the LRP5/6 and Frizzled receptors that bind Wnt. Patients with loss of function mutations in SOST show dramatically increased bone mass and sclerosteosis or van Buchem disease. Antibodies that inhibit sclerostin activity show great promise as bone anabolic agents for treating conditions associated with bone loss.

TNF-Receptors

The tumor necrosis factor (TNF) family of receptors is a large group of cytokine receptors that bind to both soluble and cell-membrane-associated ligands. One important member of this family is the receptor activator of nuclear factor kappa B (RANK) which plays a critical function in the regulation of bone physiology (see Chapter 8). These receptors consist of a trimeric complex of three single transmembrane receptors that bind to ligand. The cytoplasmic tail of many TNF-receptors (including RANK) interacts with a family of adapter molecules called TNF receptor–associated factors (TRAFs) that further activate a number of downstream pathways, the most of important of which is the activation of nuclear factor kappa B (NF-κB) (Figure 1–13). Activation of NF-κB is a central event in many if not all inflammatory responses and leads to the induction of many molecules including those involved in apoptotic, cytokine, and chemokine pathways. TNF-receptor signaling leads to the activation of NF-κB through phosphorylation of the heterotrimeric I-κB kinase (IKK) complex, which then phosphorylates the inhibitor of NF-κB (I-κB). I-κB complexes with NF-κB in the cytosol and keeps it inactive; however, when I-κB is phosphorylated it dissociates from the complex and undergoes degradation through the ubiquitin-dependent

proteasome pathway. NF-κB then translocates to the nucleus where it regulates gene transcription.

WNT/Beta Catenin

Signaling through the Wnt/β-catenin complex is critical for embryonic development and cellular homeostasis. The β-catenin
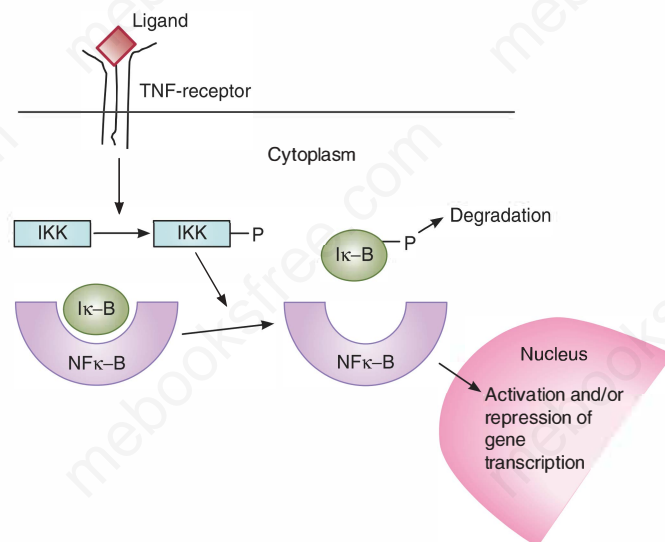
FIGURE 1–13 TNF receptor signaling. TNF binds to a trimolecular cell surface receptor which transmits a downstream signal that leads to the phosphorylation and activation of the I-κB kinase (IKK). IKK phosphorylates the inhibitor of NF-κB (I-κB) which is then tagged for degradation through a ubiquitin-dependent proteosomal pathway. This allows NF-κB to migrate to the nucleus where it can carry out activation or repression of gene transcription.

proteins are found in two main complexes: as a cadherin cellular adhesion protein-bound form in adherens junctions (Figure 1–14A), and in a complex with APC, Axin, and GSK3 as part of the Wnt signaling pathway (Figure 1–14B).

At the cell membrane, β-catenins serve to promote cell-cell adhesions by linking cadherins to actin filaments (see Figure 1–14A). Mutations in β-catenins appear to play a critical role in the progression towards cancer by regulating the degradation of the Wnt/β-catenin complex or by changing cellular adhesion properties. These mutations are found in colorectal, breast, ovarian, and endometrial cancers.

Beta-catenin can also be regulated by secreted Wnt proteins, providing a key link between extracellular Wnt signaling and activation of genetic developmental programs (see Figure 1–14B). In the Wnt signaling pathway, levels of β-catenins are primarily controlled by ubiquitination and proteosomal degradation. This is triggered through phosphorylation by glycogen synthase kinase 3 alpha and beta (GSK3α and β) and casein kinase I (CKI). This process is amplified by the scaffolding proteins Disheveled and Axin. These components assemble to create a large complex that marks β-catenin molecules for subsequent destruction by proteasomes.

Recently, the activation of the Wnt/β-catenin pathway in osteocytes has been identified as a critical regulator of anabolic bone activity, while activation of the same pathway in osteoblasts may be responsible for decreased bone resorption. As described previously, mutations in the Wnt receptor LRP5/6 can also change osteoblast anabolic function, contributing to high bone mass (activating mutations in LRP5) and diseases such as osteoporosis-pseudoglioma syndrome (loss of function mutations in LRP5). These features make the Wnt/β-catenin pathway particularly intriguing as a therapeutic target for a wide spectrum of disorders.
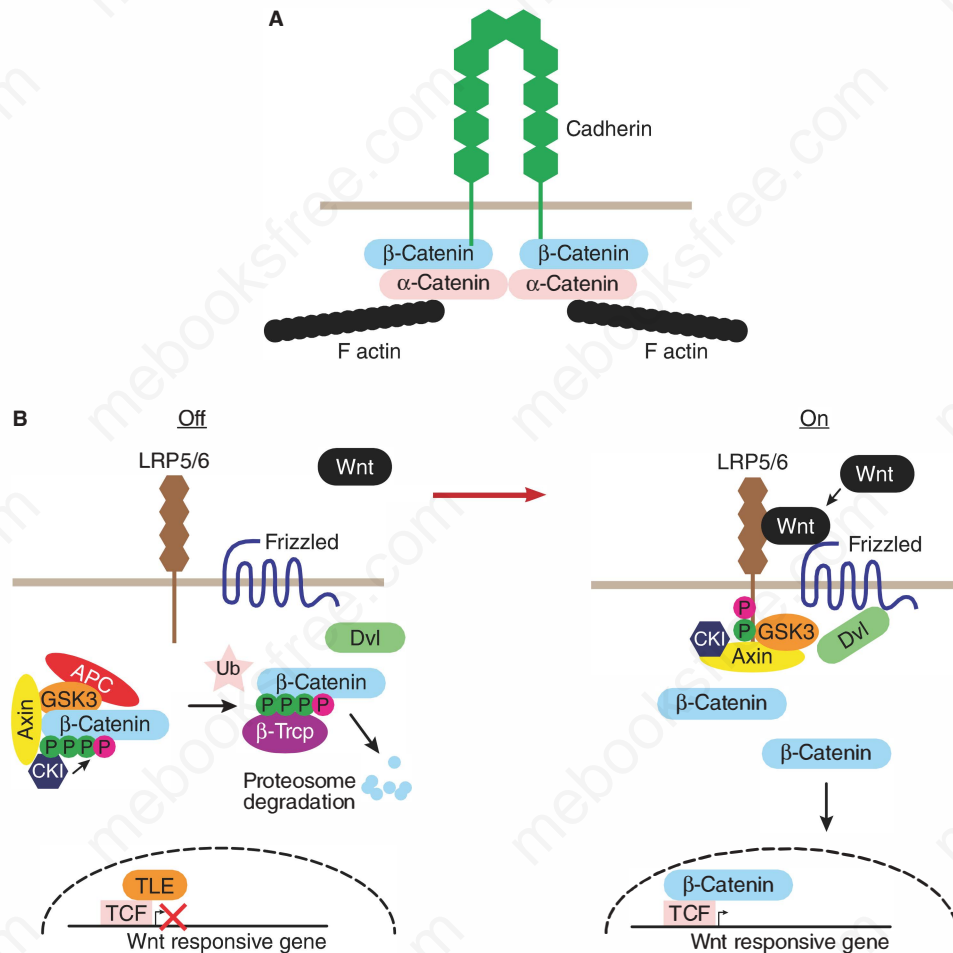
FIGURE 1–14 Beta-catenin signaling. β-catenin proteins are found in two main signaling complexes: as a cadherin cellular adhesion protein-bound form in adherens junctions **(A)** that promotes cell-cell adhesion and as part of the Wnt signaling pathway in a complex associated with APC, Axin, and GSK3 **(B)**. When the pathway is inactive, the β-catenin complex forms and CKI adds the 4th phosphate, thus activating ubiquitination and subsequent proteasome degradation of β-catenin. Activation of the Wnt signaling receptors Frizzled and LRP5/6 allows unphosphorylated β-catenin to enter the nucleus and activate Wnt responsive genes.

GUANYLYL CYCLASE–LINKED RECEPTORS

Activation of guanylyl cyclase–dependent signaling cascades can occur through two independent mechanisms. The first involves activation of the soluble guanylyl cyclase, a heme-containing enzyme that is activated by the gas nitric oxide (NO) generated in the same or neighboring cells. NO is produced by the heme-containing enzyme nitric oxide synthase. NO synthase exists as three different isozymes in selected body tissues. Constitutive forms of NO synthase (NOS) are produced in endothelial (NOS-3) and neuronal (NOS-1) cells. The endothelial enzyme possesses binding sites for FAD and FMN as well as calcium and appears to require calcium for optimal activity. Endothelial NOS activity is regulated by myristoylation and palmitoylation, which anchor the enzyme in the membrane compartment, and by phosphorylation. Agents such as bradykinin and acetylcholine, which interact with receptors on the surface of endothelial cells and increase intracellular calcium levels, trigger an increase in constitutive NO synthase activity with consequent generation of NO and activation of soluble guanylyl cyclase activity in neighboring vascular smooth muscle cells (Figure 1–15). Thus, in this instance, the cGMP-dependent vasodilatory activity of acetylcholine requires sequential waves of signaling activity in two different cell types to realize the ultimate physiologic effect.

The inducible (i) form of NO synthase (NOS-2) is found predominantly in inflammatory cells of the immune system, although it has also been reported to be present in smooth muscle cells of the vascular wall. Unlike the endothelial form of NO synthase, expression of iNO synthase is low in the basal state. Treatment of cells with a variety of cytokines triggers an increase in new iNO synthase synthesis (hence, the inducible component of iNO synthase activity), probably through activation of specific *cis* elements in the iNO synthase promoter. Thus, hormones, cytokines, or growth factors with the capacity for induction of iNO synthase activity may direct at least a portion of their signaling activity through a cGMP-dependent pathway.

A third mechanism for increasing cGMP levels within target cells involves the activation of particulate guanylyl cyclases (Figure 1–16). From an endocrine standpoint, this involves predominantly the natriuretic peptide receptors (NPR). NPR-A is a single-transmembrane-domain receptor (about 130 kDa) with a large extracellular domain that provides ligand recognition and binding. This is followed by a hydrophobic transmembrane domain and a large intracellular domain that harbors the signaling function. The amino terminal portion of this intracellular region contains an ATP-binding kinase homology domain (KHD) that is involved in regulating cyclase activity, whereas the carboxyl terminal domain contains the catalytic core of the particulate guanylyl cyclase. It is believed that association of ligand with the
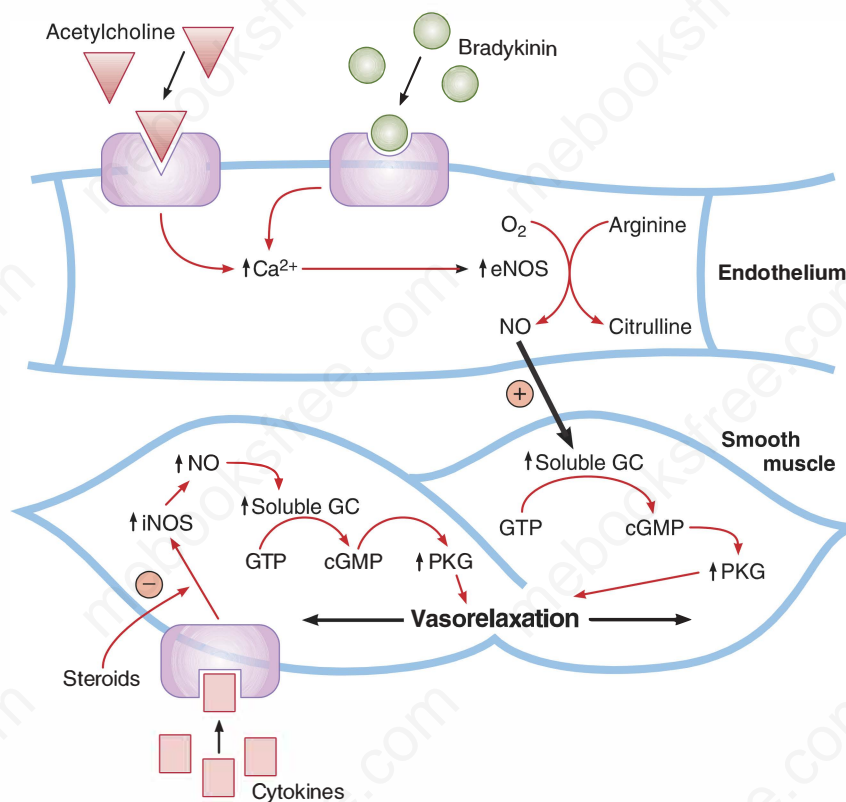
FIGURE 1–15 Signaling through the endothelial (e) and inducible (i) nitric oxide synthases (NOS) in the vascular wall. Activation of eNOS in the endothelial cell or iNOS in the vascular smooth muscle cell leads to an increase in NO and stimulation of soluble guanylyl cyclase (GC) activity. Subsequent elevations in cGMP activate cGMP-dependent protein kinase (PKG) and promote vasorelaxation.
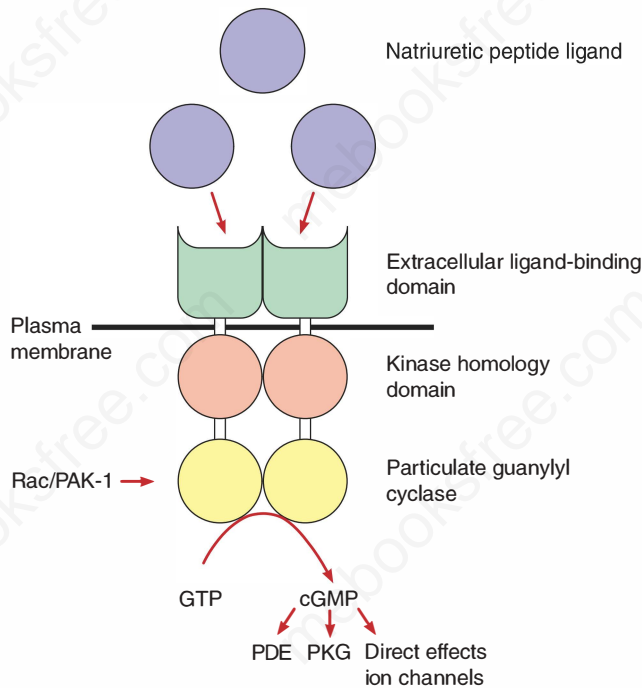
FIGURE 1–16 Signaling by particulate guanylyl cyclase. Ligand (ANP or BNP for the type A natriuretic peptide receptor or CNP for the type B receptor) associates with the extracellular domain of the receptor. This effects a change in the receptor that reduces affinity for ligand and alters conformation of the KHD which, in turn, reverses tonic inhibition of particulate guanylyl cyclase activity at the carboxy terminal portion of the receptor. Rac and PAK-1 are thought to activate the receptor through a more direct interaction with the guanylyl cyclase domain. Increased cyclase activity increases cellular cGMP levels which increase protein kinase G (PKG) activity through a mechanism similar to that described for cAMP (see Figure 1–6), regulates phosphodiesterase (PDE) activity, and alters the ion transport properties of the target cell.

extracellular domain leads to a conformational change in the receptor that arrests the tonic inhibitory control of the kinase-like domain and permits activation of guanylyl cyclase activity. Recent studies suggest that the small GTPase Rac can directly activate the catalytic domain through its effector kinase PAK (p21-activated kinase). NPR-B, the product of a separate gene, has a similar topology and a relatively high level of sequence homology to the NPR-A gene product; however, while NPR-A responds predominantly to the cardiac atrial natriuretic peptide (ANP) and brain natriuretic peptide (BNP), NPR-B is activated by the C-type NP (CNP), a peptide found in the central nervous system, endothelium, and reproductive tissues but not in the heart. Thus, segregated expression of the ligand and its cognate receptor convey a high level of response specificity to these two systems despite the fact that they share a common final effector mechanism. Noteworthy, both NPR-A and NPR-B require a high degree of phosphorylation in the KHD to retain sensitivity to agonist. Dephosphorylation, perhaps through agonist-triggered exposure of the phosphoserine residues to regulatory phosphatase activity in the target cell, results in receptor desensitization. In humans,

homozygous mutation of the *NPR-B* genes results in acromesomelic dysplasia Maroteaux type (AMDM), a rare form of autosomal recessive, short-limbed dwarfism.

NUCLEAR ACTION OF PEPTIDE HORMONES

Although the initial targets of peptide hormone receptor signaling appear to be confined to the cytoplasm, it is clear that these receptors can also have profound effects on nuclear transcriptional activity. They accomplish this through the same mechanisms they use to regulate enzymatic activity in the cytoplasmic compartment (eg, through activation of kinases and phosphatases) via second-messengers that transmit information into the nucleus. In this case, however, the ultimate targets are transcription factors that govern the expression of target genes. Examples include hormonal activation of c-Jun and c-Fos nuclear transcription factors that make up the heterodimeric AP-1 complex. This complex has been shown to alter the expression of a wide variety of eukaryotic genes through association with a specific recognition element, termed the phorbol ester (TPA) response element (TRE), present within the DNA sequence of their respective promoters. Other growth factor receptors that use the MAPK-dependent signaling mechanism appear to target the serum response factor (SRF) and its associated ternary complex proteins. Post-translational modification of these transcription factors is believed to amplify the signal that traffics from this complex, when associated with the cognate serum response element (SRE), to the core transcriptional apparatus. The cAMP-dependent activation of protein kinase A results in the phosphorylation of a nuclear protein CREB (cAMP response element–binding protein) at Ser^{119}, an event that results in enhanced transcriptional activity of closely positioned promoters. The latter requires the participation of an intermediate CREB-binding protein (CBP). CBP is a co-activator molecule that functionally tethers CREB to proteins of the core transcriptional machinery. Interestingly, CBP may also play a similar role in nuclear receptor (NR) signaling (see Nuclear Receptors). Several recent studies have provided evidence suggesting that a number of peptide hormones and growth factors may bind to high-affinity receptors in the cell nucleus. The role these receptors play—if any—in contributing to the signaling profile of these peptides remains undefined.

NUCLEAR RECEPTORS

The NRs, which include those for the glucocorticoids, mineralocorticoids, androgens, progesterone, estrogens, thyroid hormone, and vitamin D, differ from the receptors of the surface membrane described earlier in that they are soluble receptors with a proclivity for using transcriptional regulation as a means of promoting their biologic effects. Thus, although some receptors are compartmentalized in the cytoplasm (eg, glucocorticoid receptor), whereas others are confined to the nucleus (eg, thyroid hormone receptor), they all operate within the nuclear chromatin to initiate the

signaling cascade. These receptors can be grouped into two major subtypes based on shared structural and functional properties. The first, the steroid receptor family, includes the prototypical glucocorticoid receptor (GR) and the receptors for mineralocorticoids (MR), androgens (AR), and progesterone (PR). The second, the thyroid receptor family, includes the thyroid hormone receptor (TR) and the receptors for estrogen (ER), retinoic acid (RAR and RXR), and vitamin D (VDR) as well as the peroxisome proliferator–activated receptor (PPAR). There are also so-called orphan receptors that bear structural homology to members of the extended NR family. For many of these, the "ligand" is unknown, and their functional roles in the regulation of gene expression have yet to be determined. Others have been linked to specific ligands. For example, PPARs bind to fatty acids, FXR with isoprenoids and bile acids, LXR with oxysterols, PXR with steroid metabolites, CAR with xenobiotics, and BXR with folic acid derivatives. In total, there are more than 70 nuclear receptors and orphan receptors in humans.

STEROID RECEPTOR FAMILY

Steroid receptors (ie, GR, MR, AR, and PR), under basal conditions, exist as cytoplasmic, multimeric complexes that include the heat shock proteins (HSPs) 90, 70, and 56 and immunophilins of the FK506 family (FKBP51 and FKBP52). ER, although demonstrating similar association with HSP, is largely confined to the nuclear compartment. Association of the steroid ligand with the receptor results in dissociation of the HSP. This in turn exposes a nuclear translocation signal previously buried in the receptor structure (near DNA binding and hinge regions, see later) and initiates transport of the receptor to the nucleus, where it associates with the hormone response element (Figure 1–17).

Each of the family members has been cloned and sequenced, and crystallographic structures have been obtained for many of them. Consequently, we know a great deal about their structure
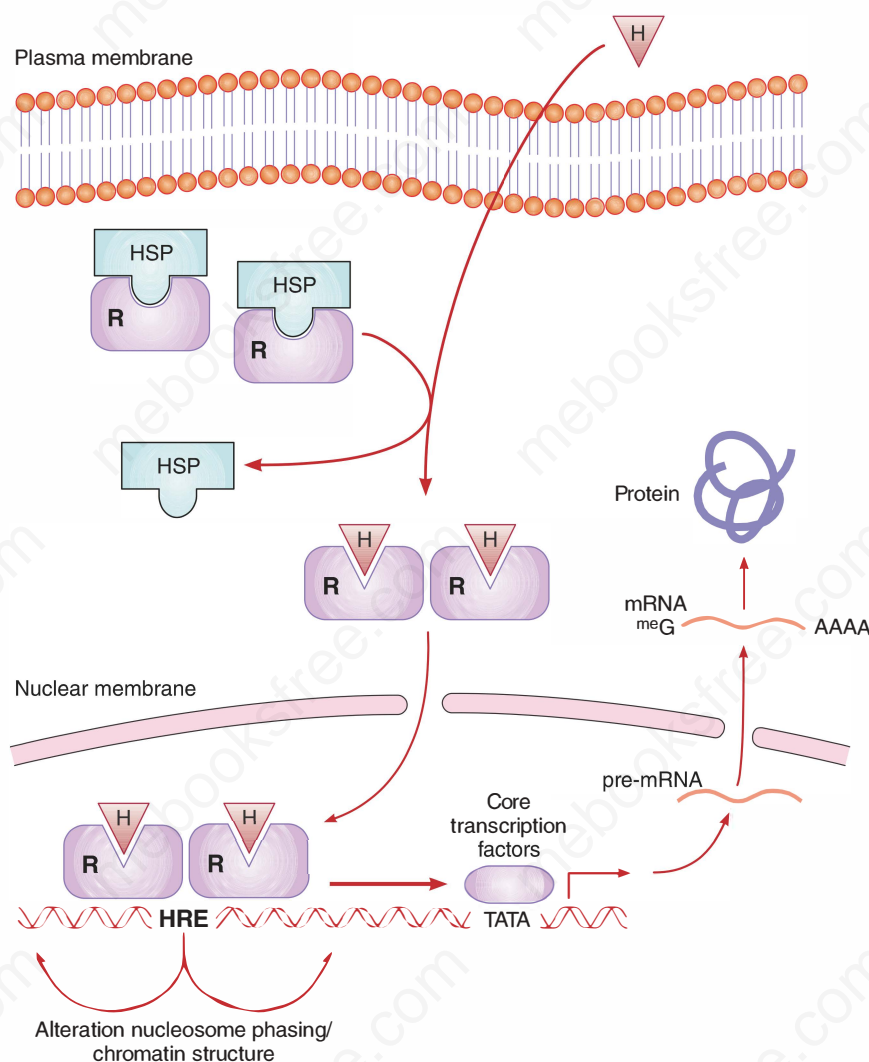
FIGURE 1–17 Signaling through the steroid receptor complex. Similar mechanisms are employed by members of the *TR* gene family, though most of the latter are concentrated in the nuclear compartment and are not associated with the HSP complex prior to binding ligand (^meG, methyl guanosine).
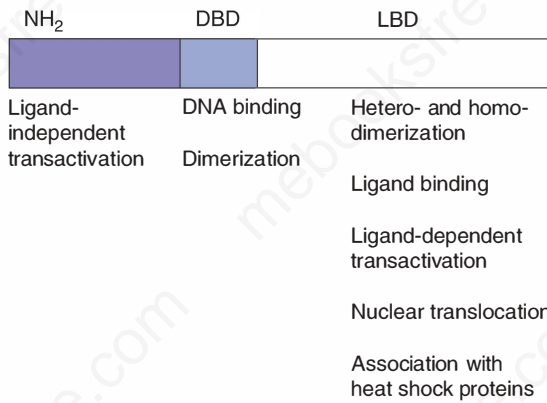
FIGURE 1–18 Structural schematic of a representative steroid receptor molecule. Separate designations are given to the amino terminal (NH$_2$), DNA-binding (DBD), and ligand-binding (LBD) domains. Functional activity associated with each of these individual domains, as determined by mutagenesis studies, is indicated under each individual domain.

and function (Figure 1–18). Each has an extended amino terminal domain of varying length and limited sequence homology to other family members. In at least some receptors, this region, which has been termed AF-1, is believed to participate in the transactivation function through which the individual receptors promote increased gene transcription. Significant variability in the length of the amino terminal regions of the different receptors suggests potential differences in their respective mechanisms for transcriptional regulation. In fact, different AF-1 domains display differential affinity for interaction with coactivator proteins (see later). The amino terminal is followed by a basic region that has a high degree of

sequence homology in both the steroid and thyroid receptor gene families. This basic region encodes two zinc finger motifs (Figure 1–19) that have been shown to establish contacts in the major groove of the cognate DNA recognition element (see the discussion later). Based on crystallographic data collected for the DNA-binding region of the GR, we know that the amino acid sequence lying between the first and second fingers (ie, recognition helix) is responsible for establishing specific contacts with the DNA. The second finger provides the stabilizing contacts that increase the affinity of the receptor for DNA. The DNA-binding region also harbors amino acid residues that contribute to the dimerization of monomers contiguously arrayed on the DNA recognition element. Following the basic region is the carboxyl terminal domain of the protein. This domain is responsible for binding of the relevant ligand, receptor dimerization or heterodimerization, and association with the HSPs. It also contributes to the ligand-dependent transactivation function (incorporated in a subdomain termed AF-2) that drives transcriptional activity. Interestingly, in selected cases, nonligands have been shown to be capable of activating steroid receptors. Dopamine activates the progesterone receptor and increases PR-dependent transcriptional activity, probably through a phosphorylation event, which elicits a conformational change similar to that produced by the association of the receptor with progesterone.

The DNA-binding regions of these receptors contact DNA through a canonical HRE, which is described in Table 1–3. Interestingly, each receptor in the individual subfamily binds to the same recognition element with high affinity. Thus, specificity of hormone action must be established either by contextual DNA sequence lying outside the recognition element or by other, nonreceptor DNA-protein interactions positioned in close proximity
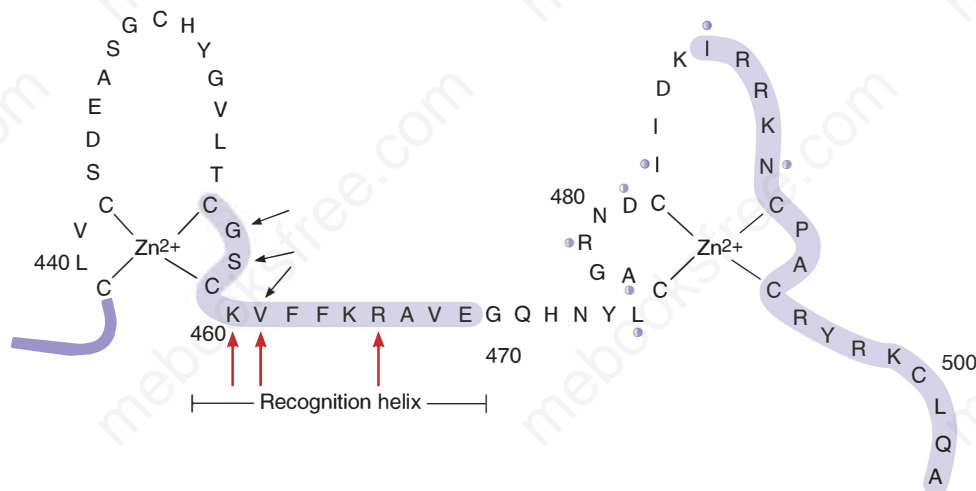
FIGURE 1–19 Schema of the two zinc fingers, together with coordinated zinc ion, which make up the DNA-binding domain of the GR (amino acids are numbered relative to the full-length receptor). Shaded regions denote two alpha helical structures which are oriented perpendicularly to one another in the receptor molecule. The first of these, the recognition helix, makes contact with bases in the major groove of the DNA. Red arrows identify amino acids which contact specific bases in the glucocorticoid response element (GRE). Black arrows identify amino acids that confer specificity for the GRE; selective substitutions at these positions can shift receptor specificity to other response elements. Dots identify amino acids making specific contacts with the phosphate backbone of DNA. (Reproduced with permission from Luisi BF, Xu WX, Otwinowski Z, et al. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*. 1991 Aug 8;352(6335):497-505.)

TABLE 1–3 DNA recognition elements for major classes of nuclear hormone receptors.[a]

Element	Recognition Sequence	Receptor
HRE	AGAACANNNTGTTCT	Glucocorticoid Mineralocorticoid Progesterone Androgen
ERE	AGGTCANNNTGACCT	Estrogen
TRE	AGGTCA(N)$_n$AGGTCA	Vitamin D Thyroid hormone Retinoic acids PPAR ligands

[a]Elements represent consensus sequences selected to emphasize the modular nature of the half-sites and their capacity for palindrome generation. Sequences read in the 5′ to 3′ direction. N denotes a spacer nucleoside (either A, G, C, or T). Half-sites are identified by the overlying arrows. The TRE is arrayed as a direct repeat but may also exist as a palindrome or an inverted palindrome. A variable number of spacer nucleotides are positioned between the two direct repeats, depending on the type of hormone receptor. Three, four, and five nucleosides (ie, n = 3, 4, or 5) are preferred for binding of the VDR, TR, or RAR, respectively.

to the element. Interestingly, the GR, as well as some other NRs (eg, ER), are capable of binding to DNA sequence lacking the classic HRE. Originally described in the mouse proliferin gene promoter, these composite elements associate with heterologous complexes containing GR, as well as with components of the AP-1 transcription factor complex (ie, c-Jun and c-Fos). They display unique regulatory activity at the level of contiguously positioned promoters. One such composite element, for example, directs very specific transcriptional effects depending on whether the GR or the MR is included in the complex.

Several steroids, particularly the glucocorticoids and estrogens, have been reported to have independent effects on the stability of target gene transcripts. At this point, it is unclear what role the hormone receptors play in this process and whether transcript stabilization is tied mechanistically to the enhancement of transcriptional activity.

THYROID RECEPTOR FAMILY

The thyroid receptor family includes the TR, RAR, RXR, ER, PPAR, and VDR. They share a high degree of homology to the proto-oncogene c-erbA and high affinity for a common DNA recognition site (see Table 1–3). With the exception of the ER, they do not associate with the HSPs, and they are constitutively bound to chromatin in the cell nucleus. Specificity of binding for each of the individual receptors is, once again, probably conferred by a contextual sequence surrounding this element, the orientation of the elements (eg, direct repeats or inverted repeats or palindromes), the polarity (ie, 5′ in contrast to 3′ position on two successive repeats), and the number and nature of the spacing nucleotides separating the repeats.

The ER binds to its RE as a homodimer, whereas the VDR, RAR, RXR, and TR prefer binding as heterodimers. The nature of the heterodimeric partners has provided some intriguing insights into

the biology of these receptors. The most prevalent TR-associated partners appear to be the RXRs. These latter receptors, which as homodimers form high-affinity associations with 9-cis-retinoic acid, also form heterodimeric complexes in the unliganded state with VDR and RAR. In the individual cases in which it has been examined, heterodimerization with RXR amplifies both the DNA binding and the functional activity of these other receptors. Thus, the ability to form such heterodimeric complexes may add significantly to the flexibility and potency of these hormone receptor systems in regulating gene expression. Interestingly, the positioning (5′ vs 3′) of the participant proteins on the RE is important in determining the functional outcome of the association. In most of those situations linked to transcriptional activation, RXR seems to prefer the upstream (5′) position in the dimeric complex. Thus, diversity of response is engendered by the selection of recognition elements (eg, monomeric vs dimeric vs oligomeric sites) and by the choice and positioning of the dimeric partner (eg, homodimer vs heterodimer) where applicable.

The crystallographic structures of the LBDs of several members of the TR family have been described. These include the dimeric unliganded RXRα, monomeric liganded RARγ, monomeric liganded TRα, dimeric agonist (ie, estradiol)—and antagonist (ie, raloxifene)—liganded ERα, liganded VDR, and liganded PPARγ. A composite LBD displays a common folding pattern with 12 alpha helices (numbered by convention H1-H12) and a conserved beta turn. Some variability exists in that there is no H2 in RARγ and a short H2′ helix is present in PPARγ, but the overall structural configuration is preserved. The dimeric interface is formed through interaction of amino acids located in helices 7 to 10, with the strongest influence exerted by H10. These interactions appear to be important for both homodimeric as well as heterodimeric interactions. Binding of ligand has been shown to occur through what has been termed a "mousetrap" mechanism. In the unliganded state, H12, which contains the carboxyl terminal activation domain AF-2, is displaced away from the ligand-binding pocket (Figure 1–20). Association of agonist ligand (eg, estradiol in the case of the ER) with the hydrophobic core of the receptor leads to a repositioning of H12 over the ligand-binding cavity, where it stabilizes receptor–ligand interactions and closes the "mousetrap." Binding of an antagonist ligand such as raloxifene, which because of its structure engenders steric hindrance in the ligand-binding pocket, prevents closure of H12 into the normal agonist position. Instead, H12 folds into an alternative location between H4 and H3, a conformation that suppresses the activation function of the receptor (discussed later). Crystal structure of a full-length nuclear receptor pair (in this case PPAR gamma and RXR alpha) shows that the PPAR gamma protein dominates its heterodimeric partner, dictating the overall structure of the complex.

The mechanistic underpinnings of transcriptional regulation by the NRs have been partially elucidated (Figure 1–21). In the unliganded or antagonist-bound state, the receptor dimers are preferentially associated with a macromolecular complex containing the repressor proteins N-CoR or SMRT, a transcriptional corepressor Sin3, and a histone deacetylase RPD3. N-CoR and SMRT each use two independent receptor interaction domains

A

H4

H5

H3

B
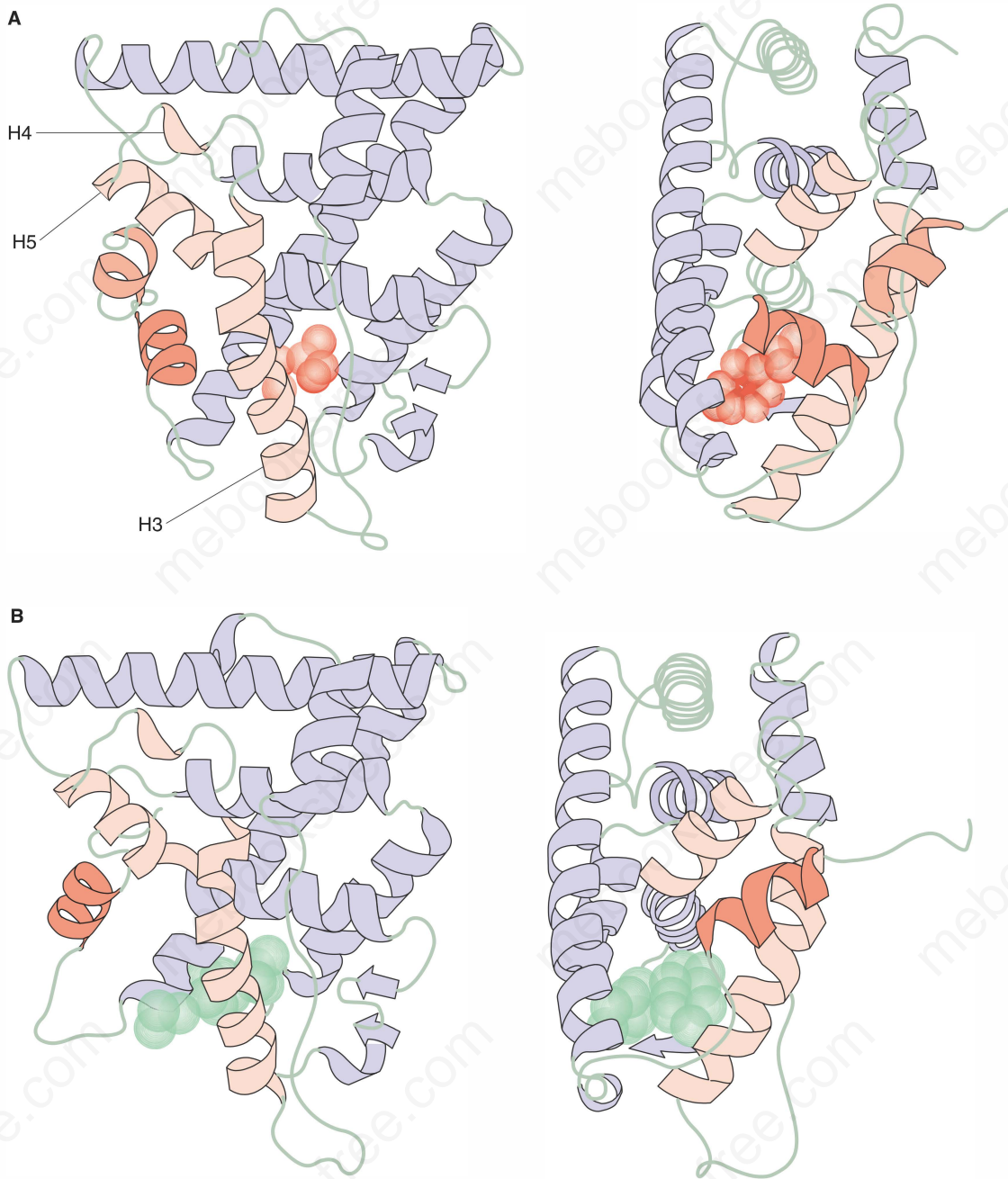
FIGURE 1–20   Three-dimensional structures for the agonist- and antagonist-occupied ERα LBD. **(A)** Orthogonal views of the agonist diethylstilbestrol-ERα LBD-NR Box II peptide complex. Coactivator peptide and LBD are presented as ribbon diagrams. Peptide is colored medium orange, helix (H) 12 (ERα residues 538-546) is colored dark orange. Helices 3, 4, and 5 are colored light pink. Diethylstilbestrol is depicted in red in a space-filling format. **(B)** Orthogonal views of the antagonist 4-hydroxytamoxifen-ERα LBD complex. Color scheme is the same as in panel A. 4-Hydroxytamoxifen is shown in green in a space-filling format. Note that NR Box II is absent in this structure. **(C)** Schematic representation of the mechanism underlying agonist-dependent activation of nuclear hormone receptor (see next page.) In the presence of agonist, helix 12 (the terminal helix in the LBD) folds across the ligand-binding pocket, stabilizing ligand–receptor interaction and promoting a conformation conducive for coactivator association. In the presence of antagonist, steric hindrance precludes folding of helix 12 across the ligand-binding pocket. Instead, it positions itself in the region typically occupied by the coactivator, thereby blocking the activation function of the receptor. (Reproduced with permission from Shiau AK, Barstad D, Loria PM, et al. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*. 1998 Dec 23;95(7):927-937.)

C
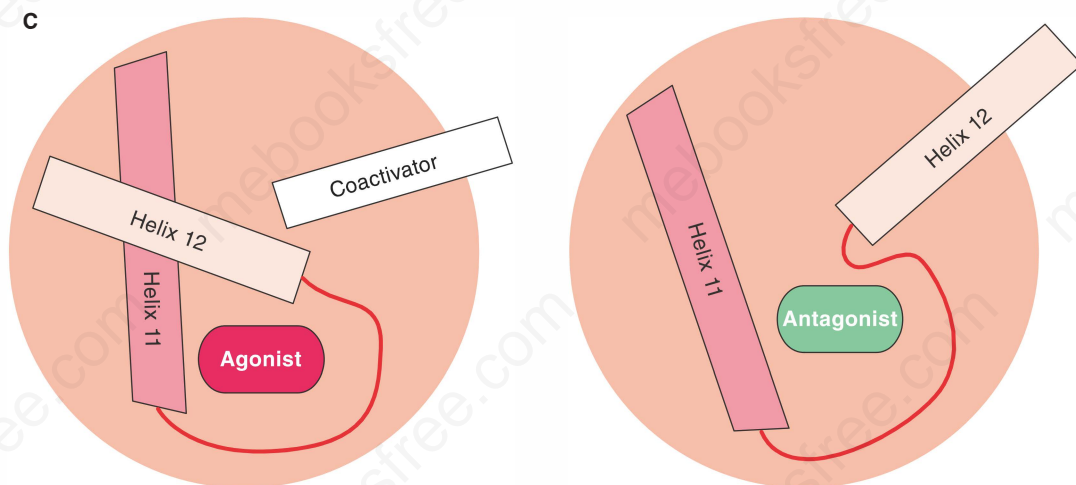
FIGURE 1–20 (*Continued*)
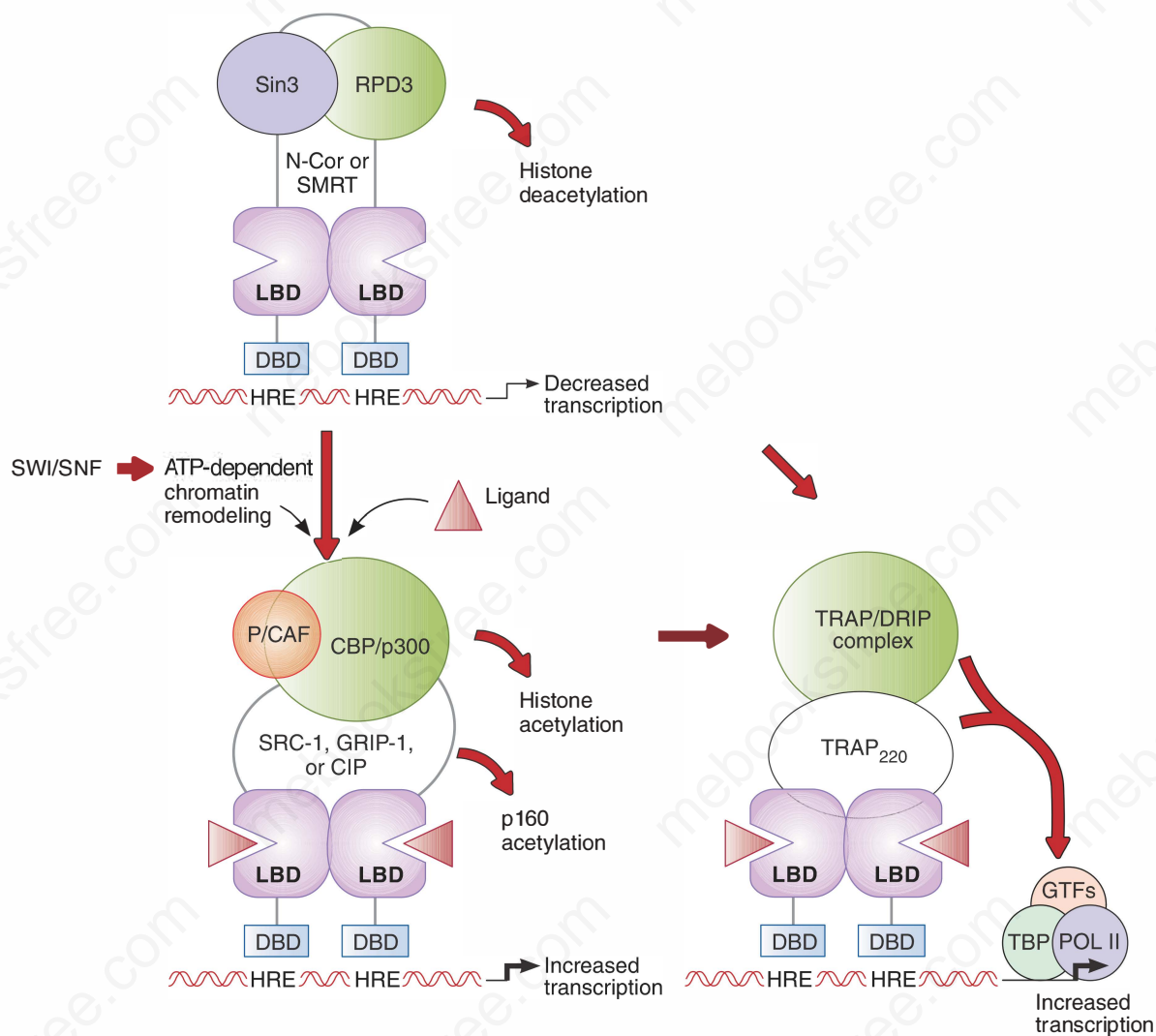
FIGURE 1–21 Interaction of corepressor (top) versus coactivator (bottom) molecules with the LBD of a representative nuclear receptor (see text for details). The temporal order of p160 versus DRIP/TRAP–binding remains undetermined.

(IDs) to associate with the NRs (one repressor; two receptors). Each ID contains an amino acid sequence (L/IXXI/VI, where I = isoleucine, L = leucine, V = valine, X = any amino acid), which interacts with helices 4, 5, and 6 of the NR LBD. Histone acetylation is typically associated with activation of gene transcription (presumably reflecting decompaction of chromatin surrounding the transcriptional unit), so the presence of histone deacetylase activity in the complex is thought to promote a transcriptionally quiescent state. Addition of ligand leads to a change in receptor conformation that no longer favors interaction with the corepressor (a shift in position of helix 12 in the LBD prevents corepressor interaction and promotes coactivator assembly into the complex) and promotes both ATP-dependent chromatin remodeling and assembly of an activator complex containing p160 coactivator proteins (eg, SRC-1, GRIP-1, or P/CIP) and, secondarily, the CREB-binding protein (CBP) and the histone acetylase P/CAF. The net accrual of histone acetylase activity (CBP and P/CIP as well as P/CAF possess acetylase activity) leads to acetylation of chromatin proteins (eg, histones) as well as components of the core transcriptional machinery, resulting in chromatin decompaction and a net increase in transcriptional activity. Interaction of the NRs with the coactivators in this complex takes place through LXXLL motifs (where L = leucine and X = any amino acid) present in the coactivator proteins. Each coactivator may have several of these motifs, which preferentially associate with different NRs, other transcription factors, or other coactivators. This allows for a degree of selectivity in terms of which regulatory proteins are incorporated in the complex. Notably, a recent structural analysis showed that a 13-amino-acid peptide, containing an LXXLL motif from the GRIP-1 protein, interacts with TRβ through a hydrophobic cleft generated by helices 3, 4, and 12 (including AF-2) in the receptor protein. This is the same cleft that is occupied by helix 12, which harbors an LXXLL motif, in the raloxifene-bound ERα. This suggests that the antagonist in the latter instance acquires its activation-blocking properties by repositioning helix 12 in a manner that leads to displacement of the coactivator protein from this groove (see discussion earlier). SRC also interacts with the AF-1 domain, suggesting a potential mechanism for maximizing synergistic activity between AF-1 and AF-2 domains in the receptors. The importance of SRC1 in thyroid hormone action is underscored by the finding that SRC1 knockout mice display significant thyroid hormone resistance in peripheral tissues.

CBP is thought to function as a pivotal component of the NR regulatory complex. While the p160 class of coactivators interact directly with the NRs (primary coactivator), CBP associates primarily with the p160 coactivators, thereby establishing an indirect link to the receptors (secondary coactivator). CBP also has the ability to establish weaker, primary associations with the receptor proteins. As noted earlier, CBP functions as a central integrator of transcriptional regulatory signals from multiple pathways, including the cAMP-dependent activation of the transcription factor CREB. Recent evidence suggests that an additional level of regulatory control may be involved in selectively amplifying NR-dependent transcriptional activity. An enzyme called coactivator-associated arginine methyltransferase 1 (CARM1) associates with CBP and methylates the protein. This results in a reduction in CREB-dependent gene activation and, secondarily, an increase in NR-dependent gene transcription. This switching mechanism effectively refocuses the transcriptional machinery on expression of NR-dependent gene expression. CARM1 also interacts with all three members of the SRC family. It binds to the AD2 domain at the carboxy terminus of the p160 coactivators (by contrast, CBP associates with the AD1 domain located in the midregion of the molecule). CARM1 in this context promotes methylation of the p160 coactivator and dissociation of the CARM1-SRC complex from NR-associated DNA, effectively clearing the promoter of the transcriptional regulatory complex.

Another family of coactivator complexes has also been identified as playing an important role in NR signaling. The human TR-associated protein (TRAP) and vitamin D receptor–interacting protein (DRIP) complexes are the best characterized to date. These complexes, which contain in the neighborhood of 25 individual proteins, are thought to serve as a functional bridge between the liganded NR bound to DNA and the general transcription factors (GTFs) (eg, TBP, TFIIB, RNA polymerase II, and TAFs) involved in formation of the preinitiation complex, a function previously assigned to the Mediator complex in yeast. The TRAP220 subunit appears to establish the relevant contacts, through LXXLL motifs, with the NRs in promoting this assembly. Their role vis-à-vis the p160 coactivators alluded to earlier remains undefined; however, it has been suggested that they succeed the p160 coactivator complex in binding with liganded NRs positioned on target gene promoters, establish the requisite structural and functional connections with the core transcriptional machinery, and initiate mRNA synthesis. It has also been suggested that acetylation of one of the key NR-binding motifs (LXXLL) on the SRC coactivator by CBP leads to dissociation of SRC from the nuclear receptors, thereby allowing access for assembly of the TRAP/DRIP complex (see Figure 1–21).

Nuclear receptors can be post-translationally modified through acetylation, ubiquitination, and phosphorylation. Modifications are common in the amino terminal AF-1 domain where they can affect the affinity for selected coactivator proteins and, thereby, modulate receptor activity. Phosphorylation, which is the most extensively studied of these modifications, can affect virtually all nuclear receptors. Phosphorylation by kinases associated with the general transcriptional machinery (eg, cyclin-dependent kinase 7) or major intracellular signaling pathways (eg, Akt) may serve to recruit coactivators into the complex, thereby facilitating transcriptional regulation. Coactivator molecules or other factors recruited into the multisubunit transcription complexes may possess kinase, ubiquitin ligase, ATPase, methyltransferase, or acetylase activity which can target either nuclear receptors or other coregulators. Thus, in addition to their ability to provide the scaffolding for complex assembly, they also possess the enzymatic capacity to fine tune the transcriptional activity of the complex. Coactivators themselves can undergo post-translational modification. Phosphorylation of the SRC coactivators, for example, can increase coactivator affinity for selected nuclear steroid receptors (eg, AR, PR, and ER), as well the p300 and CBP proteins, thereby increasing the stability of the activation complex. In addition to

their effects on transcription, coactivators have been suggested to be involved in transcript elongation, RNA splicing, and mRNA transport.

Recent studies have shown that a significant portion of ER DNA-binding sites in MCF7 cells (a breast cancer cell line) are topographically linked to, or in some cases directly overlap with, those for the forkhead box transcription factor A1 (FOX A1). FOX A1 has been referred to as a pioneer factor that aids in positioning ER (or AR) at specific sites on DNA. FOX A1 is recruited to specific epigenetic markers in chromatin; in the case of MCF7 cells this is a lysine 4 methylation on histone 3. This, in turn, directs ER binding to nearby regions on the DNA. Similar findings have been reported with the AR, albeit with different epigenetic chromatin markers.

Although the GR is encoded by a single gene, there are several translation start sites and the primary gene transcript can be alternatively spliced to generate a number of variants, some of which (eg, GR β) may exert dominant negative activity when paired with the predominant receptor (eg, GR α). GR beta is abundantly expressed in neutrophils and epithelial cells where it may be responsible for suppressing glucocorticoid activity. TR is encoded by two genes (α and β). TR α1 and TR β1 appear to be the dominant forms of TR in the body. Although considerable overlap exists in their tissue distribution, TR α1 is enriched in skeletal muscle, brown fat, and the central nervous system, and TR β1 is found in the liver, kidney, and central nervous system. They are believed to signal most of the developmental and thermogenic effects of thyroid hormone in the whole animal. TR β2, a splice variant of the TR β gene, is found in the rodent pituitary gland, where it may subserve a specific regulatory function (eg, control of TSH secretion). TR α2, an alternatively spliced product of the TR α gene, lacks the HBD at the carboxyl terminal of the molecule and thus is not a true thyroid hormone receptor. Under certain experimental conditions, TR α2 can block the activity of other members of the TR family, but its physiologic role, if one exists, remains undefined.

Similar heterogeneity exists in the retinoid receptor family. There are three isoforms for both the RXR and RAR. Collectively, these receptors are thought to play an important role in morphogenesis, but the function of the individual isoforms remains only partially understood. There are two ER isoforms. ERα appears to signal the bulk of traditional estrogenic activity. ERβ is found in a variety of different tissues and possesses antiproliferative activity that may serve to dampen the effects of liganded ER.

NONGENOMIC EFFECTS OF THE STEROID HORMONES

Although steroids exert most of their biologic activity through direct genomic effects, there are several lines of evidence suggesting that this does not provide a complete picture of steroid hormone action. There are several examples that, for kinetic or experimental reasons, do not fit the classic paradigm associated with a transcriptional regulatory mechanism. Included within

this group are the rapid suppression of ACTH secretion following steroid administration, the modulation of oocyte maturation and neuronal excitability by progesterone, the stimulation of endothelial nitric oxide synthase (through interaction of ERα with the p85 subunit of PI-3K) by estrogen, the inhibition of type II deiodinase and stimulation of mitochondrial oxygen consumption by thyroid hormone, and regulation of calcium channel function by 1,25-$(OH)_2$ vitamin D. Recent studies have demonstrated the presence of conventional sex steroid, vitamin D, and thyroid hormone receptors on membranes of target cells. Attachment of these to the membrane requires palmitoylation of internal cysteine residues that leads to association of the membrane-bound receptor with caveolae. These caveolae traffic the receptors to signaling complexes or "rafts" within the membrane. The PR displays a novel ability to interact with the SH3 domains of the Src family tyrosine kinases, thereby accessing the Ras/Raf/MEK-1/ERK signaling pathway (discussed earlier). Although we still do not completely understand the mechanisms underlying these nongenomic effects, their potential importance in mediating steroid or thyroid hormone action may, in selected instances, approach that of their more conventional genomic counterparts.

Neurosteroids represent another class of nontraditional hormonal agonists with unique biologic activity. Some of these are native steroids (eg, progesterone), whereas others are conjugated derivatives or metabolites of the native steroids (eg, dihydroprogesterone). These agonists have been identified in the central nervous system and in some instances shown to have potent biologic activity. It is believed that they operate through interaction with the receptor for γ-aminobutyric acid, a molecule that increases neuronal membrane conductance to chloride ion. This has the net effect of hyperpolarizing the cellular membrane and suppressing neuronal excitability. Interactions that promote receptor activity would be predicted to produce sedative-hypnotic effects in the whole animal, whereas inhibitory interactions would be expected to lead to a state of central nervous system excitation.

STEROID AND THYROID HORMONE RECEPTOR RESISTANCE SYNDROMES

Heritable defects in these receptors have been linked to the pathophysiology of a number of hormone-resistance syndromes. These syndromes are characterized by a clinical phenotype suggesting hormone deficiency, by elevated levels of the circulating hormone ligand, and increased (or inappropriately detectable) levels of the relevant trophic regulatory hormone (eg, ACTH, TSH, FSH, or LH). Point mutations in the zinc fingers of the DNA-binding domain as well the LBD of the VDR lead to a form of vitamin D–dependent rickets (type II) characterized by typical rachitic bone lesions, secondary hyperparathyroidism, and alopecia. It is inherited as an autosomal recessive disorder. Molecular defects scattered along the full length of the AR (>400 to date), although concentrated in the LBD, have been linked to syndromes

characterized by varying degrees of androgen resistance ranging from infertility to full-blown testicular feminization syndrome. Clinical severity, in this case, is thought to be related to the severity of the functional impairment that the mutation imposes on the receptor. Because the AR is located on the X chromosome, these disorders are inherited in an X-linked fashion.

Defects in the GR are less common, perhaps reflecting the life-threatening nature of derangements in this system. However, mutations have been identified that impact negatively on receptor function. Clinical presentations in these cases have been dominated by signs and symptoms referable to glucocorticoid deficiency (eg, fatigue, asthenia) and adrenal androgen (eg, hirsutism and sexual precocity) and mineralocorticoid (eg, low renin hypertension) overproduction. This presumably results from defective steroid-mediated suppression of ACTH secretion and adrenal hyperplasia as the former rises in an attempt to restore glucocorticoid activity at the periphery.

Resistance to thyroid hormone has been linked to a large number of mutations scattered along the full length of the β form of the receptor, although, once again, there is a concentration of mutations in the LBD, particularly along the rim of the coactivator-binding pocket. The clinical presentation of thyroid hormone resistance resulting from a defective TR beta extends from the more typical mild attention-deficit syndromes to full-blown hypothyroidism with impaired growth. Different target tissues harboring the mutant receptors display variable sensitivity to thyroid hormone, with some tissues (eg, pituitary) displaying profound resistance and others (eg, heart) responding in a fashion suggesting hyperstimulation with thyroid hormone (ie, thyrotoxicosis). The latter effects (eg, tachycardia) may reflect the predominance of the normal α isoform, as opposed to defective β TR isoform, in the target tissue (eg, heart). Mutation of the alpha form of TR was described recently. It leads to significant growth and developmental retardation, skeletal dysplasia, and severe constipation. These syndromes are rather unique in that they are inherited as autosomal dominant disorders, presumably reflecting the ability of the mutated receptors to interfere with receptors produced from the normal allele, either by binding to the RE with higher affinity than the wild-type receptors and precluding access of the latter to target genes or by forming inactive heterodimers with the wild-type receptor proteins (ie, dominant negative mutant receptor). Mutations in the thyroid receptor can also result in agonist-dependent association with transcriptional repressors (eg, N-CoR and SMRT) resulting in a reduction in transcriptional activity. A number of patients with thyroid hormone resistance have been described that lack mutations in the TR. In two cases, the defect has been assigned to defects in the transport and intracellular metabolism of thyroid hormone.

Defects in the ER are rare, perhaps reflecting the critical role estrogens play in regulating lipoprotein metabolism. However, one male patient who harbors a mutation within the LBD of the ER has been described. His clinical presentation was characterized by infertility as well as osteopenia, suggesting important roles for estrogens in the maintenance of spermatogenesis as well as bone growth even in male subjects.

A novel mutation in the PPARγ receptor leads to a phenotype that resembles that of patients with the metabolic syndrome (ie, severe insulin resistance, diabetes mellitus, and hypertension). A syndrome of mineralocorticoid resistance, or pseudohypoaldosteronism, has been described in a number of independent kindreds. Pseudohypoaldosteronism type I is characterized by neonatal renal salt wasting, dehydration, hypotension, hyperkalemia, and hyperchloremic metabolic acidosis despite the presence of elevated aldosterone levels. Heterozygous mutations in the MR are responsible for a milder form of the disease which is inherited in an autosomal dominant pattern with abnormalities largely confined to the kidney. A more severe form of the disease with generalized systemic involvement is inherited in an autosomal recessive pattern and appears to be due to loss-of-function mutations in genes encoding subunits of the amiloride-sensitive epithelial sodium channel. Of equivalent interest is the recent identification of an activating mineralocorticoid receptor mutation (Ser_{810}-to-Leu_{810}). This mutation results in severe early-onset hypertension that is markedly exacerbated by pregnancy. The mutation leads to constitutive activation of the MR and alters the specificity of ligand binding such that traditional MR antagonists, like progesterone, function as partial agonists. This latter property presumably accounts for the dramatic increase in blood pressure during pregnancy. Genetic deficiency of coactivators have also been linked to human disease. Expression levels of PPARγ coactivator alpha (PGC-1α) have been linked to metabolic disorders like diabetes mellitus and obesity. Mutation of the CBP gene leads to the Rubenstein-Taybi syndrome characterized by defective development and impaired long-term memory formation.

REFERENCES

Hormones and Hormone Action

Funder JW. Receptors, hummingbirds and refrigerators. *News Physiol Sci*. 1987;2:231.

G Protein Receptors

Barki-Harrington L, Rockman HA. β-Arrestins: multifunctional cellular mediators. *Physiology*. 2008;23:17. [PMID: 18268361]

Lania AG, Mantovani G, Spada A. Mechanisms of disease: mutations of G proteins and G-protein-coupled receptors in endocrine diseases. *Nat Clin Pract Endocrinol Metab*. 2006;2:681. [PMID: 17143315]

Métayé T, Gibelin H, Perdrisot R, Kraimps JL. Pathophysiological roles of G-protein-coupled receptor kinases. *Cell Signal*. 2005;17:917. [PMID: 15894165]

Schöneberg T, Schulz A, Biebermann H, Hermsdorf T, Römpler H, Sangkuhl K. Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacol Ther*. 2004;104:173. [PMID: 1556674]

Weinstein LS, Chen M, Xie T, Liu J. Genetic diseases associated with heterotrimeric G proteins. *Trends Pharmacol Sci*. 2006;27:260. [PMID: 16600389]

Non-Genomic Steroid Receptors

Thomas P, Pang Y, Filardo EJ, Dong J. Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. *Endocrinology*. 2005;146:624. [PMID: 15539556]

Yin H, Chu A, Li W, et al. Lipid G protein-coupled receptor ligand identification using beta-arrestin PathHunter assay. *J Biol Chem*. 2009;284:12328. [PMID: 19286662]

Wnt/β-Catenin and Sclerostin

Balemans W, Ebeling M, Patel N, et al. Increased bone density in sclerosteosis is due to the deficiency of a novel secreted protein (SOST). *Hum Mol Genet.* 2001;10:537. [PMID: 11181578]

MacDonald BG, Tamai K, He X. Wnt/β-Catenin signaling: components, mechanisms, and diseases. *Developmental Cell.* 2009;17:9. [PMID: 19619488]

Sharifi M, Ereifej L, Lewiecki EM. Sclerostin and skeletal health. *Rev Endocr Metab Disord.* 2015;16:149. [PMID: 25669441]

Tu X, Delgado-Calle J, Condon KW, et al. Osteocytes mediate the anabolic actions of canonical Wnt/β-catenin signaling in bone. *Proc Natl Acad Sci USA.* 2015;112:E478. [PMID: 25605937]

G Protein Effectors

Cazzolli R, Shemon AN, Fang MQ, Hughes WE. Phospholipid signaling through phospholipase D and phosphatidic acid. *IUBMB Life.* 2006;58:457. [PMID: 16916782]

Kiselyov K, Shin DM, Muallem S. Signaling specificity in GPCR-dependent Ca^{2+} signaling. *Cell Signal.* 2003;15:243. [PMID: 12531423]

Luttrell LM. Transmembrane signaling by G protein-coupled receptors. *Methods Mol Biol.* 2006;332:3. [PMID: 16878684]

Werry TD, Sexton PM, Christopoulos A. "Ins and outs" of seven-transmembrane receptor signaling to ERK. *Trend Endocrinol Metab.* 2005;16:26. [PMID: 15620546]

Woehler A, Ponimaskin EG. G protein-mediated signaling: same receptor, multiple effectors. *Curr Mol Pharmacol.* 2009;2:237. [PMID: 20021461]

Tyrosine Kinase–Coupled and Cytokine Receptors

Belfiore A, Frasca F, Pandini G, Sciacca L, Vigneri R. Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. *Endocr Rev.* 2009;30:586. [PMID: 19752219]

Hupfeld CJ, Olefsky JM. Regulation of receptor tyrosine kinase signaling by GRKs and beta-arrestins. *Annu Rev Physiol.* 2007;69:561. [PMID: 17002595]

Kikani CK, Dong LQ, Liu F. "New"-clear functions of PDK1: beyond a master kinase in the cytosol? *J Cell Biochem.* 2005;96:1157. [PMID: 16187290]

LeRoith D, Yakar S. Mechanisms of disease: metabolic effects of growth hormone and insulin-like growth factor 1. *Nat Clin Pract Endocrinol Metab.* 2007;3:302. [PMID: 17315038]

Guanylyl Cyclase-Linked Receptors

Dudzinski DM, Igarashi J, Greif D, Michel T. The regulation and pharmacology of endothelial nitric oxide synthases. *Annu Rev Pharmacol Toxicol.* 2006;46:235. [PMID: 16402905]

Gardner DG, Chen S, Glenn DJ, Grigsby CL. Molecular biology of the natriuretic peptide system: implications for physiology and hypertension. *Hypertension.* 2007;49:419. [PMID: 17283251]

Kuhn M. Function and dysfunction of mammalian membrane guanylyl cyclase receptors: lessons from genetic mouse models and implications for human disease. *Handb Exp Pharmacol.* 2009;191:47. [PMID: 19089325]

Mutchler SM, Straub AC. Compartmentalized nitric oxide signaling in the resistance vasculature. *Nitric Oxide.* 2015;49:8. [PMID: 26028569]

Nuclear Receptors

Chrousos GP, Kino T. Glucocorticoid signaling in the cell. Expanding clinical implications to complex human behavioral and somatic disorders. *Ann NY Acad Sci.* 2009;1179:153. [PMID: 19906238]

Cornell W, Nam K. Steroid hormone binding receptors: application of homology modeling, induced fit docking, and molecular dynamics to study structure-function relationships. *Curr Top Med Chem.* 2009;9:844. [PMID: 19754398]

Dasgupta S, Lonard DM, O'Malley BW. Nuclear receptor coactivators: master regulators of human health and disease. *Annu Rev Med.* 2014;65:279. [PMID: 24111892]

Dumitrescu A, Refetoff S. The syndromes of reduced sensitivity to thyroid hormone. *Biochem Biophys Acta.* 2013;1830:3987. [PMID: 22986150]

Ito M, Roeder RG. The TRAP/SMCC/mediator complex and thyroid hormone receptor function. *Trends Endocrinol Metab.* 2001;12:127. [PMID: 11306338]

Levin ER. Extranuclear steroid receptors are essential for steroid hormone action. *Annu Rev Med.* 2015;66:271. [PMID: 25587652]

Lonard DM, O'Malley B. Nuclear receptor coregulators: modulators of pathology and therapeutic targets. *Nat Rev Endocrinol.* 2012;8:598. [PMID: 22733267]

Nicolaides NC, Charmandari E. Chrousos syndrome: from molecular pathogenesis to therapeutic management. *Eur J Clin Invest.* 2015;45:504. [PMID: 25715669]

Oakley R, Cidlowski JA. The biology of the glucocorticoid receptor: new signaling mechanism in health and disease. *J Allergy Clin Immunol.* 2013;132:1033. [PMID: 24084075]

Ordonez-Moran P, Munoz A. Nuclear receptors: genomic and non-genomic effects converge. *Cell Cycle.* 2009;8:1675. [PMID: 19448403]

Rochel N, Moras D. Ligand binding domain of vitamin D receptors. *Curr Top Med Chem.* 2006;6:1229. [PMID: 16848737]

Rousseau GG. Fifty years ago: the quest for steroid hormone receptors. *Molec Cell Endocrinol.* 2013;375:10. [PMID: 23684885]

Simons SS Jr, Edwards DP, Kumar R. Minireview: dynamic structures of nuclear hormone receptors: new promises and challenges. *J Mol Endocrinol.* 2014;28:173. [PMID: 24284822]

Trevino LS, Weigel NL. Phosphorylation: a fundamental regulator of steroid receptor action. *Trends Endocrinol Metab.* 2013;24:515. [PMID: 23838532]

Yang J, Young MJ. The mineralocorticoid receptor and its coregulators. *J Mol Endocrinol.* 2009;43:53. [PMID: 19617444]

Endocrine Autoimmunity

2

Juan Carlos Jaume, MD

AADC	Aromatic L-amino acid decarboxylase	**IPEX**	Immunodeficiency, polyendocrinopathy, and enteropathy, x-linked
ACA	Antibodies recognizing the adrenal cortex	**LFA**	Lymphocyte function-associated antigen
ADCC	Antibody-dependent cell-mediated cytotoxicity	**MHC**	Major histocompatibility complex
AICD	Activation-induced cell death	**NALP1**	NACHT leucine-rich-repeat protein 1
AIRE	Autoimmune regulator gene	**NALP5**	NACHT leucine-rich-repeat protein 5
APECED	Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy	**NK**	Natural killer (cells)
APS	Autoimmune polyendocrine syndrome	**NOD**	Nonobese diabetic (mice)
BB	Bio breeding	**SCA**	Steroid-producing cell antibodies
BCR	B-cell receptor	**SCID**	Spontaneous combined immunodeficiency
cAMP	Cyclic adenosine monophosphate	**TAP**	Transporter associated with antigen processing
CaSR	Calcium-sensing receptor	**TBI**	Thyrotropin-binding inhibition
CD	Cluster of differentiation	**TCR**	T-cell receptor
CTLA	Cytotoxic T lymphocyte antigen	**TD**	Thymus-dependent
DPT	Diabetes Prevention Trial	**Tg**	Thyroglobulin
FOXP3	Forkhead box P3	**TI**	Thymus-independent
GABA	Gamma-aminobutyric acid	**TNF**	Tumor necrosis factor
GAD	Glutamic acid decarboxylase	**TPO**	Thyroperoxidase
HLA	Human leukocyte antigen	**TSH**	Thyroid-stimulating hormone
IA-2	Islet cell antigen-2 (tyrosine phosphatase)	**TSH-R**	Thyrotropin receptor
IFN	Interferon	**TSI**	Thyroid-stimulating immunoglobulin
IL	Interleukin	**VNTR**	Variable number of tandem repeats

Epidemiologic analysis of a large population reported that about 1 of 30 (3.2%) people in the United States are currently affected by autoimmune diseases. Graves disease, type 1 diabetes, pernicious anemia, rheumatoid arthritis, Hashimoto thyroiditis, and vitiligo are the most prevalent such conditions, accounting for 93% of affected individuals. A more global approach at calculating prevalence led to a corrected estimate between 7.6% and 9.4% of the world's population as affected by autoimmune diseases (2.5 in 30 people worldwide).

These autoimmune diseases have traditionally been looked upon as forming a spectrum. At one end are found organ-specific diseases with organ-specific targets. Hashimoto thyroiditis is an example in which a specific lesion affects the thyroid (lymphocytic infiltration, destruction of follicular cells) and autoantibodies are produced with absolute specificity for thyroid proteins. At the other end of the spectrum are the systemic autoimmune diseases, broadly belonging to the class of rheumatologic disorders. Systemic lupus erythematosus is an example of a disease characterized by widespread pathologic changes and a collection of autoantibodies to DNA and other nuclear constituents of all cells. Many organ-specific autoimmune diseases are autoimmune endocrinopathies. Furthermore, most endocrine glands are subject to autoimmune attack including the adrenals (autoimmune Addison disease), gonads (autoimmune oophoritis), pancreas (type 1 diabetes), pituitary (autoimmune hypophysitis), and thyroid (autoimmune thyroid disease) (Table 2–1).

TABLE 2–1 Some autoimmune endocrinopathies, antigens, and autoantibodies.

Disease	Gland	Autoantigen	Autoantibody
Autoimmune (lymphocytic) hypophysitis Granulomatous hypophysitis	Pituitary	Pituitary cytosolic protein	Anti-pituitary
Graves disease	Thyroid	TSH-R, TPO	TSI, TBII, anti-TPO
Hashimoto thyroiditis	Thyroid	TPO, Tg	Anti-TPO, anti-Tg
Autoimmune (idiopathic) hypoparathyroidism	Parathyroid	CaSR, NALP5, IFN alpha, IFN omega, other cytokine antibodies	Antiparathyroid
Type 1 diabetes mellitus	Pancreas (β cells)	GAD65, IA-2, pre-pro-insulin	Anti-GAD, anti-IA-2 (ICA), anti-insulin
Type B insulin resistance with acanthosis nigricans (rare)	Adipocytes, muscle cells	Insulin receptor	Insulin receptor blocking
Autoimmune Addison disease (autoimmune adrenal failure)	Adrenal	21-Hydroxylase 17α-Hydroxylase P450scc	Anti-21-hydroxylase (ACA) Anti-17α-hydroxylase and anti-P450scc (SCA)
Autoimmune oophoritis (premature ovarian failure)	Ovaries	Not yet identified unequivocally, 17α-hydroxylase P450scc	Also SCA in association with adrenal insufficiency
Autoimmune orchitis/Male infertility (some forms)	Testes	Sperm	Anti-sperm

By far the most common autoimmune endocrine diseases are those involving the thyroid and the endocrine pancreas. When the target is the thyroid gland and the clinical manifestation is hypothyroidism (Hashimoto thyroiditis), the prevalence is about 1% to 2%. When the manifestation is hyperthyroidism (Graves disease), the prevalence is about 0.4%. Both thyroid autoimmune disorders affect women preferentially. When the targets of the autoimmune response are the β cells of the pancreas, the clinical presentation is type 1 diabetes. The prevalence of type 1 diabetes is close to that of Graves disease (0.2%-0.5%); however, it has no gender predilection.

Basic immunologic concepts as they apply to clinical autoimmune endocrine diseases as sole entities and as polyendocrine failure syndromes are reviewed in this chapter.

BASIC IMMUNE COMPONENTS AND MECHANISMS

The immune system is constantly confronted with a variety of molecules and recognizes them as either self or foreign. The adaptive immune system has evolved to recognize virtually any foreign molecule, either in existence or yet to come. The repertoire of immune recognition molecules randomly formed by gene rearrangements is not limited by the genetic information encoded in the genome (Figure 2–1). As a result, an enormously wide array of immune recognition molecules is acquired by the human immune system. By way of illustration, the theoretical diversity of T-cell receptors (T-cell recognition molecules) by random rearrangements reaches 10^{15}. This mechanism of rearrangement also applies to B-cell recognition molecules (ie, immunoglobulins). The random mechanism of gene rearrangement, however, produces immune recognition molecules that react with self components. Potentially dangerous immune cells carrying self-reactive recognition molecules are eliminated (negatively selected) during development of T lymphocytes in the thymus and of B

lymphocytes in the bone marrow. It appears that only immune cells which react with foreign antigen strongly and with self antigen very weakly are positively selected and comprise the peripheral immune cell repertoire. This selection mechanism of immune cells is termed "central tolerance." Self-reactive immune cells that skip central tolerance and reach the periphery are managed by other control mechanisms against autoimmunity and are either eliminated, rendered unresponsive, or suppressed ("peripheral tolerance"). Failures in these mechanisms of immunologic regulation, as proposed by Mackay and Burnet in 1964, are central features of the concept of autoimmunity.

IMMUNE RECOGNITION AND RESPONSE

T and B lymphocytes are the fundamental and predominant immune cells. T lymphocyte precursors (pre-T cells) originate in the bone marrow and migrate to the thymus, where they undergo maturation and differentiation. At early stages, they express several T-cell surface molecules but still have genomic (not rearranged) configuration of their T-cell receptors (TCRs). These pre-T cells, destined to become T cells with TCR α/β chains (T α/β cells), pass through a critical phase during which self-reactive T cells are deleted by negative selection (see T-Cell Tolerance, later in this chapter). Few pre-T cells will express other types of chains on their TCR (T γ/δ cells). T α/β cells differentiate into either mature CD4 or CD8 cells. These now mature lymphocytes migrate to T-cell areas of peripheral lymphoid organs and exert their function as helper (T_H) or cytotoxic (T_C) cells when activated.

B lymphocytes mature and differentiate in the bone marrow and then migrate to the B-cell areas of lymphoid organs. Influenced by factors derived from T_H cells previously activated by professional antigen-presenting cells (APCs) such as macrophages, some B cells differentiate to become immunoglobulin M (IgM)-producing cells (plasma cells). Most of the other activated B cells that do not
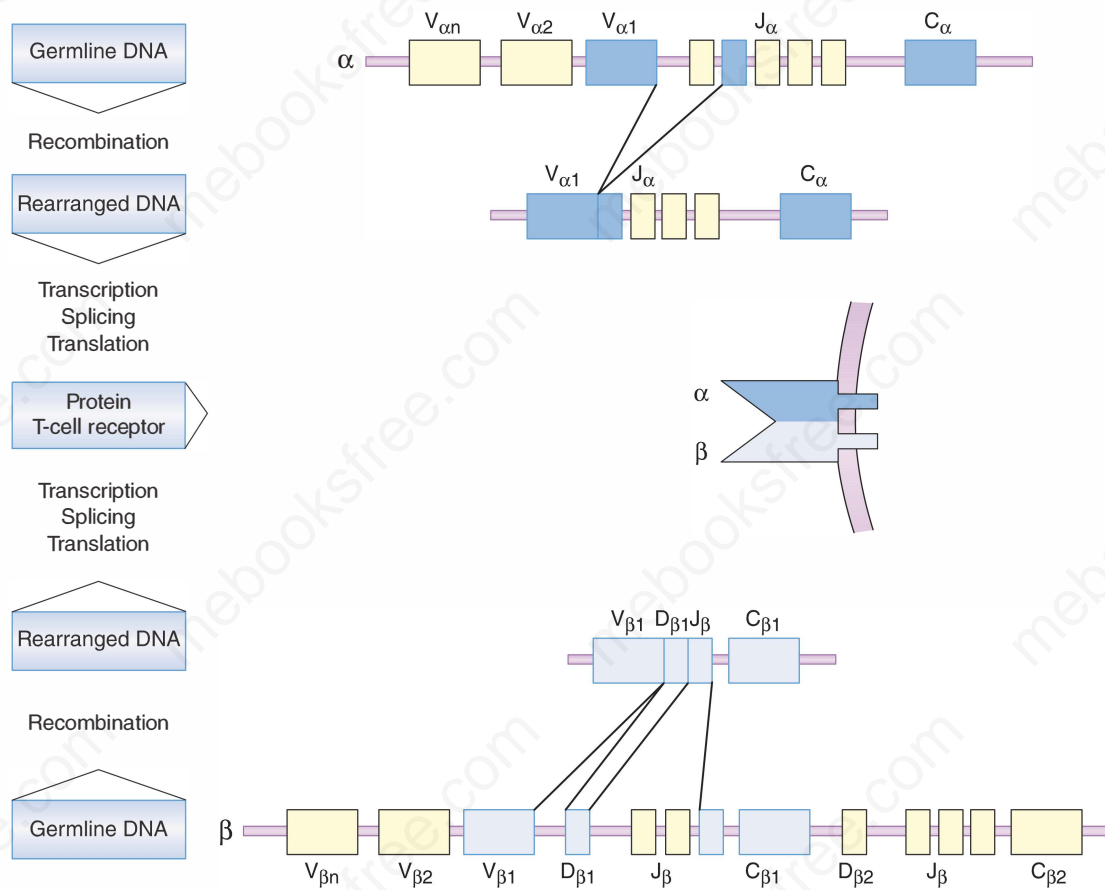
FIGURE 2-1 Rearrangement of the T-cell receptor (TCR) α and β genes to form a functionally diverse receptor. During T-cell development, the TCR α and β gene segments rearrange by somatic recombination so that one of the V_{α} segments pairs with a single J_{α} segment, and a V_{β} segment pairs with a single D_{β} and J_{β} segment. The C (constant) segments are brought together with the rearranged segments by transcription and splicing to generate the functional mRNA that will be translated into the α and β protein chains that compose the TCR.

differentiate into plasma cells revert to the resting state to become memory B cells. When memory B cells are further activated, two events occur: isotype switching (immunoglobulin class switching) and hypermutation of the immunoglobulin-variable region to further increase diversity and specificity (affinity maturation).

Activation of B cells requires recognition of the antigen as a whole, while T cells require recognition of antigenic peptides bound to major histocompatibility complex (MHC) molecules on the surfaces of professional APCs. Therefore, T-cell recognition is said to be MHC restricted.

The human MHC (human leukocyte antigen; HLA) consists of a linked set of genes encoding major glycoproteins involved in antigen presentation (Figure 2–2). The complex locates to the short arm of chromosome 6 and divides into three separate
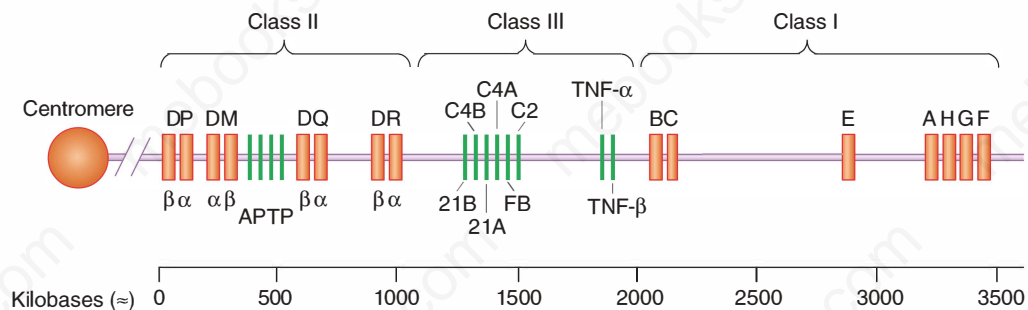
FIGURE 2-2 Gene organization of the human major histocompatibility complex or human leukocyte antigen (HLA) complex. Regions encoding the three classes of MHC proteins are on top. APTP denotes a cluster of genes within the class II region, encoding genes related to antigen processing, transport, and presentation. Class III region encodes genes unrelated to class I or class II not involved in antigen presentation (TNF-α and β, complement factors C2, C4, B, and 21-hydroxylase and others).

regions: class I, class II, and class III genes. The class I "classic" region encodes HLA-A, HLA-B, and HLA-C loci; the nonclassic or class I-related region encodes HLA-E, HLA-F, and HLA-G loci and other immunity-related genes such as CD1. The class II region (HLA-D) encodes HLA-DP, HLA-DQ, and HLA-DR loci and other genes related to antigen processing, transport, and presentation, such as transporter associated with antigen processing (TAP). The class III region encodes genes for tumor necrosis factors α and β (TNF-α and TNF-β); complement factors C2, C4, and B; and the steroidogenic enzyme 21-hydroxylase. MHC class I (classic) molecules are found on all somatic cells, whereas MHC class I nonclassic antigens are expressed only on some (eg, HLA-F on fetal liver, HLA-G on placental tissues). CD1 molecules are expressed on Langerhans cells, dendritic cells, macrophages, and B cells (all professional APCs). MHC class II molecules are exclusively expressed on these professional APCs. However, virtually all cells except mature erythrocytes can express MHC class II molecules under particular conditions (eg, stimulation with interferon-γ [IFN-γ]). As a general rule, MHC class I molecules present peptides derived from endogenous antigens that have access to cytosolic cell compartments (eg, virus) to CD8 T_C cells. On the other hand, MHC class II molecules present peptides derived from antigens internalized by endocytosis into vesicular compartments (eg, bacteria) to CD4 T_H cells. MHC class II molecules also bind peptides derived from many membrane-bound self antigens.

APCs process and present antigen in order to activate T cells utilizing MHC-peptide presentation (Figure 2–3). T cells require at least two signals to become activated. The interaction of a TCR expressed on antigen-specific T cells and the antigenic peptide–MHC complex expressed on APCs provides the first signal. The second signal is delivered primarily by the interaction between costimulatory molecules CD80 (B7.1) and CD86 (B7.2) on APCs and CD28 on T cells. These two signals induce proliferation of T cells, production of interleukin-2 (IL-2), and expression of the antiapoptotic protein Bcl-xL. T_H cells and T_C cells are effector cells that require both signals in order to become activated. However, T_C cells also need the "help" provided by T_H cells. Until recently, it was thought that T_H and T_C cells needed to interact with the same APC simultaneously and that cytokines (such as IL-2) produced by the T_H cell would then act on the T_C cell to facilitate its response. New studies suggest that the interaction between another costimulatory molecule, CD40 ligand (CD154), present on T cells, and CD40, present on APCs, may provide an alternative explanation. It appears that T_H cell–recognizing antigenic peptides presented by APCs deliver a signal through the CD154–CD40 complex that "licenses" APCs to directly stimulate T_C cells (Figure 2–4). Thus, there is no need for simultaneous interactions of T_H and T_C cells while encountering the APC. CD154–CD40 interaction also enhances expression of CD80 and CD86 as well as secretion of cytokines (IL-1, -6, -8, -10, and -12 and TNF-α).

Yet another molecule on T cells, the CD28 homolog cytotoxic T lymphocyte antigen 4 (CTLA-4 or CD152), functions to suppress T-cell responses (see Figure 2–3). CD152 is expressed at low to undetectable levels on resting T cells. It is upregulated by the
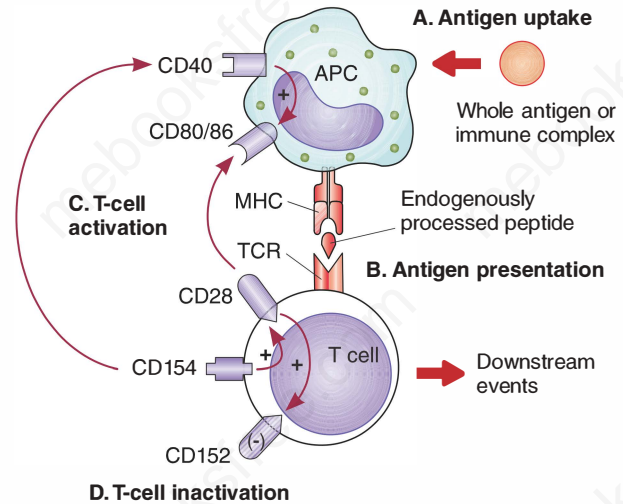
FIGURE 2–3 Antigen recognition by T cells. From top to bottom: **A. Antigen uptake**: Incorporation of antigen (via phagocytosis, pinocytosis, or FcR-mediated endocytosis of immune complex). **B. Antigen presentation**: APCs deliver an antigen-specific signal through the MHC-peptide–TCR interaction on T cells (MHC I coupled to CD8 interacts with T_C cells, MHC II coupled to CD4 interacts with T_H cells). **C. T-cell activation**: The required second signal is provided via CD80/86 (B7.1; B7.2)–CD28 that induces the expression of CD154 (CD40 L) first and CD152 (CTLA-4) later on. Binding of CD154 on T cells with CD40 on APCs enhances expression of CD80/86. The APC–CD80/86 increased expression and consequent binding of CD28 on T cells perpetuates the activation and proliferation of these effector cells (downstream events). **D. T-cell inactivation**: CD152 (expressed 48-72 hours after T-cell activation) will preferentially bind to CD80/86 on APCs because of its higher affinity, displacing CD28 and in turn suppressing T-cell activity.
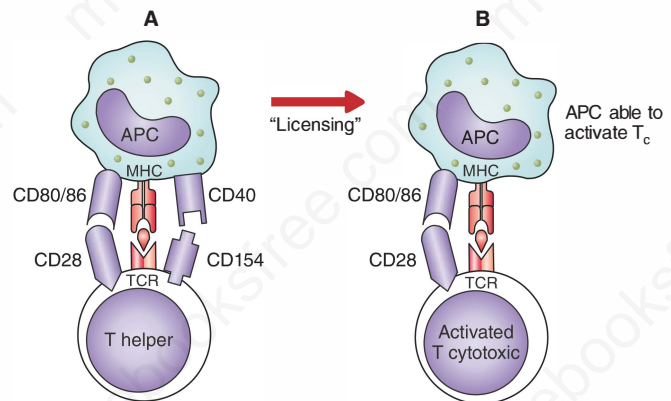
FIGURE 2–4 Licensed APCs directly activate T_C cells. According to the traditional model, T_H cells and T_C cells recognize antigen on the same APC. The APC-activated T_H cell produces IL-2, which contributes to the activation of T_C cells while in simultaneous interaction with the same APC. According to the proposed new model **(A)**, APCs are licensed to activate T_C cells by T_H or other stimuli (lipopolysaccharides, IFN-γ, viruses). APCs first interact with T_H cells. The association of CD154 (CD40 L) on the T_H cell and CD40 on the APC allows (or licenses) the latter to activate T_C cells directly **(B)**. Thus, once licensed, APCs are capable of activating cytotoxic T cells without the need of simultaneous interaction with T_H cells.

ligation of CD28 on T cells with CD80/86 on APCs, or by IL-2. CD152 and CD28 on T cells share the same counterreceptors, namely, CD80/86 on APCs. However, CD152 has a 20-fold higher affinity than CD28 for their ligands.

The integration of all these interactions may be as follows (see Figure 2–3): After processing antigen, APCs deliver an antigen-specific first signal through the MHC-peptide–TCR interaction on T cells. A second signal is provided by a costimulatory interaction of the CD80/86–CD28 complex that induces the expression of CD154 first and then CD152. Binding of CD154 on T_H cells with CD40 on APCs enhances expression of CD80/86 and licenses APCs for direct activation of T_C cells. Other inflammatory cytokines as well as lipopolysaccharides and viruses may do the same. The increased expression of APC–CD80/86 and consequent binding of CD28 on T cells then perpetuates the activation and proliferation of these effector cells. However, the expression of CD152 48 to 72 hours after T-cell activation leads to the preferential binding of this molecule to CD80/86 on APCs because of its higher affinity for CD80/86. This may displace CD28 from CD80/86 and, in turn, suppress T-cell activity. This sequence of complex events is probably simplistic relative to what nature has to offer. A new B7 family of receptors has been reported: some with positive costimulatory capacity, many with a role in down-regulating immune responses, yet others with dual functions. For some of these pathways, such as B7H3/H4 (expressed in APCs with unknown counterreceptor on T cells), very little is known, but even for those pathways that have been extensively studied—such as the CD80/86:CD28/CTLA-4 (mentioned earlier), B7:ICOS (inducible costimulator), and PD-L1/PD-L2:PD1—new insights are still being generated. The intricacies of controlling T-cell activation are enhanced by the complexity of the costimulatory pathways, such that there are multiple possible receptor–ligand interactions.

Activation and differentiation of B cells often require two signals, too. Naive B cells are triggered by antigen but may also require accessory signals that come from activated T_H cells. Some antigens can directly activate naive B cells without the need for T_H cells (eg, lipopolysaccharides from gram-negative bacteria or polymeric protein structures). The former type of B-cell activation (MHC class II–restricted T cell help) is called thymus dependent (TD). The latter type is called thymus independent (TI). T_H cells also control isotype switching and initiate somatic hypermutation of antibody variable-region genes (see Tolerance, discussed later). Interaction between CD154 on T_H cells and CD40 on B cells and the cytokines produced by T_H cells are essential for isotype switching and formation of germinal centers in peripheral lymphoid organs. The immunoglobulin isotype switching is critical for the generation of functional diversity of a humoral immune response. Somatic hypermutation (point mutations of the variable-region genes of immunoglobulins during the course of an immune response) is necessary for the affinity maturation of antibodies.

Overall, the immune response is a combination of effector mechanisms that function to eliminate pathogenic organisms. These effector mechanisms include, as **innate immunity**, phagocytosis (by macrophages, neutrophils, monocytes, and dendritic cells) and cytotoxicity (by natural killer [NK] cells); and as **adaptive immunity**, antibody-dependent complement-mediated cytotoxicity, antibody-dependent cell-mediated cytotoxicity (ADCC), cytotoxicity by T γ/δ cells that recognize heat shock proteins on target cells, and cytotoxicity by CD8 or CD4 T_C cells. CD8 and CD4 T_C cells are activated by the described recognition of specific antigenic peptides bound to class I (for CD8), class II (for CD4) MHC molecules on the APCs and classically by IL-2 from nearby activated CD4 T_H cells. These cells kill the target by either secreting cytotoxins (perforin, granzyme) or by inducing apoptosis through the Fas–FasL (Fas ligand) interaction (killer cells carrying FasL molecules activate programmed cell death in target cells expressing Fas molecules). FasL or CD95L is a type II transmembrane protein that belongs to the TNF family. The binding of FasL with its receptor induces apoptosis. FasL–Fas receptor interactions play an important role in the regulation of the immune system.

The specificity of the immune response is crucial if self-reactivity is to be avoided. In order to ensure that lymphocyte responses and the downstream effector mechanisms they control are directed exclusively against foreign antigens and not against "self" components, a number of safety-check barriers must be negotiated before autoreactive lymphocytes can differentiate and proliferate.

TOLERANCE

T-Cell Tolerance

T cells developing in the thymus (pre-T cells) are destined to become T α/β cells through rearrangement of the TCR β gene initially, followed by the TCR α gene (Figure 2–5). If unproductive rearrangements of TCR genes occur (nonfunctional TCR α or β proteins), apoptosis of these pre-T cells follows (Figure 2–5A). If functional rearrangements of TCR α and β proteins occur, cells express TCR α/β dimer and CD3 molecules at low levels on the cell surface. TCR-rearranged cells proliferate 100-fold. Positive and negative selection occurs based on the ability of the rearranged TCR α/β to recognize antigenic peptides in association with self-MHC molecules on thymic epithelial and dendritic cells. Negative selection (clonal deletion) appears to take place in the thymus medulla, where pre-T cells bearing TCRs specific for self peptides bound to self-MHC molecules are deleted. At least 97% of developing T cells undergo apoptosis within the thymus (central tolerance). Positively selected pre-T cells increase expression of TCR α/β, express either CD4 or CD8, and become mature T cells. These mature T cells exit the thymus and go to the periphery. CD4 T cells are activated in the periphery in an MHC class II–restricted fashion, while CD8 T cells are activated in an MHC class I–restricted fashion.

A differential avidity model in which the fate of T cells is determined by the intrinsic affinity of TCRs for their ligands has been advanced to explain the paradox between positive and negative selection. According to this model, T cells with high avidity for MHC-self peptide complexes would be eliminated (negative selection), whereas T cells with low avidity to MHC-self peptide
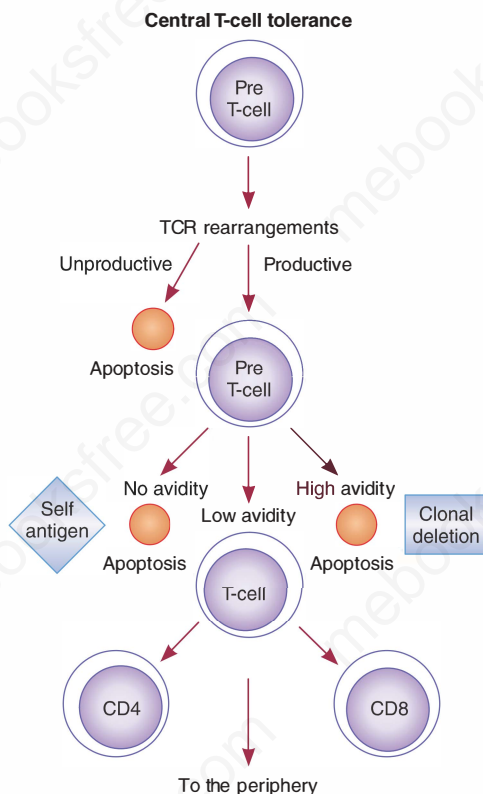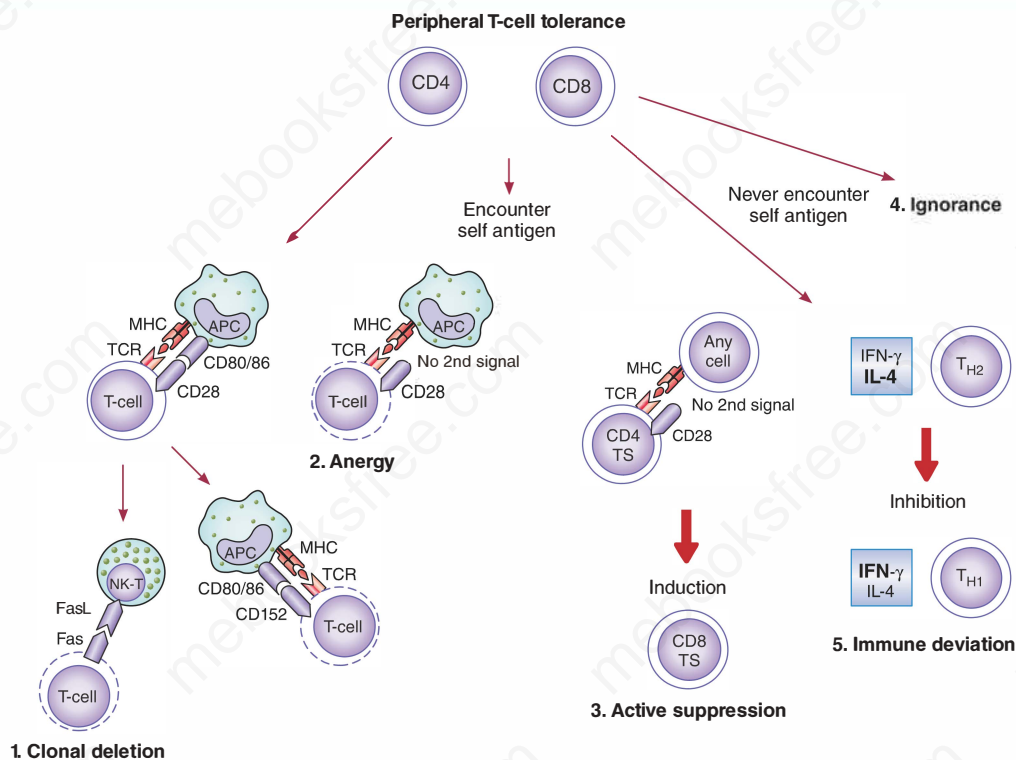
FIGURE 2–5   A. Central T-cell tolerance: Mechanisms of central tolerance (at the thymus level) are depicted. From top to bottom, pre-T cells first rearrange their TCR. Unproductive (nonfunctional) rearrangements lead to apoptosis, while productive ones engage pre-T cells in self-antigen recognition. Clonal deletion indicates elimination of cells based on their high or no avidity for self antigen (apoptosis). Surviving low-avidity cells reach the periphery as mature CD4 and CD8 cells. **B. Peripheral T-cell tolerance**: May be accomplished through any of the five depicted mechanisms. **1. Clonal deletion**: After encountering self antigen in the context of self-MHC molecules and simultaneous delivery of a second signal (CD80/86–CD28) by APCs (top left), autoreactive T cells become activated. These activated T cells express Fas molecules on their

complexes would be positively selected. If the avidity is close to zero, T cells would not be selected (for lack of effective signal to survive). The biochemical factor or factors that signal survival (low avidity of TCR binding) versus apoptosis (triggered by high avidity interactions) have yet to be found.

Costimulatory interactions between CD28 and CD80/86 and between CD154, CD40, and adhesion molecules, such as lymphocyte function-associated antigen-1 (LFA-1), are also involved in preferential deletion of self-reactive T cells in the medullary region of the thymus. It is known that negative selection is not 100% effective and that some potentially autoreactive T cells do escape to the periphery. Not all self peptides would be presented to pre-T cells during their development in the thymus. Self peptides derived from secluded proteins (ie, intracytoplasmic enzymes) only timely expressed after rigid regulatory control (ie, puberty) in endocrine glands are believed to be a likely source. Therefore, the peripheral immune system must maintain tolerance through complementary control mechanisms.

"Peripheral tolerance" (Figure 2–5B) may be maintained by the induction of unresponsiveness to self antigen (**anergy**) or by the induction of regulatory T cells (T regs), such as suppressor T cells. Peripheral **clonal deletion** (apoptosis) of autoreactive T cells that have escaped from the thymus may play an important role in limiting rapidly expanding responses, but there are many examples where autoreactive T cells persist. Some autoreactive T cells may never encounter the self antigen because it may be sequestered from the immune system (**ignorance**). Lastly, **immune deviation**, whereby noninflammatory T_{H2} responses suppress an autoreactive inflammatory T_{H1} response, inducing peripheral tolerance, deserves further discussion. T_{H1} cells, which regulate cell-mediated responses, secrete IFN-γ and small amounts of IL-4. In contrast, T_{H2} cells, which provide help for antibody production, secrete abundant IL-4 and little IFN-γ. A prevailing concept in human autoimmunity is that T_{H1} responses are believed to dominate. It has been shown in animal models that induction of T_{H2} responses ameliorates T_{H1} responses. Hence, unbalanced T_{H1} immune deviation may lead to a breakage of peripheral tolerance. However, evidence to the contrary exists in some endocrinopathies. (See autoimmune response in the section on Autoimmune Aspects of Thyroid Disease, later in the chapter.)

Clonal deletion and **anergy** occur through apoptosis at the site of activation or after passage through the liver. High antigen dose and chronic stimulation induce peripheral elimination of both CD4 and CD8 T cells. Activated T cells express Fas molecules on their surfaces but are resistant to FasL-mediated apoptosis because of the simultaneous expression of Bcl-xL (apoptosis-resistance molecules), induced by CD28 ligation during activation (see Immune Recognition and Response, earlier in the chapter). Several days after activation, when Bcl-xL has declined, CD4 cells become susceptible to Fas-mediated apoptosis (activation-induced cell death; AICD). A similar mechanism via p75 TNF receptor has been shown for CD8 cells. Therefore, autoreactive T cells might be deleted by apoptosis induced by chronic stimulation with self antigens, present abundantly in the periphery. However, autoreactive T cells specific for very rare self antigens may be difficult to eliminate.

Anergy also results from the lack of a second costimulatory signal. When nonhematopoietic cells stimulated by IFN-γ present antigen in an MHC class II–restricted fashion (as thyrocytes do in AITD), autoreactive T cells may be rendered unresponsive because of the absence of a CD28–CD80/86-mediated signal (nonhematopoietic cells do not express CD80/86 as professional APCs do). However, even if the two signals are provided, anergy may result from the lack of T_H cell-originated cytokines (IL-2, -4, -7, etc). It has also been shown that *in vivo* T-cell anergy may be induced by CD80/86–CD152 interaction (see also Immune Recognition and Response, discussed earlier).

T-cell **active suppression** is considered to be a major regulatory mechanism of peripheral tolerance; however, its mode of action is still under study. As mentioned earlier, nonhematopoietic cells stimulated by IFN-γ present antigen in an MHC class II–restricted fashion to T cells and render them anergic. These nonhematopoietic cells (nonprofessional APCs) may also present to CD4 T-suppressor cells (T_S, also known as CD4 + CD25 + FOXP3 + T regs). Before becoming unresponsive, these cells may induce specific CD8 T suppressor (T_S) cells. In turn, these CD8 T_S cells may regulate (via T-cell-suppressor factors or cytotoxicity) autoreactive T cells (see also Figure 2–5B).

B-Cell Tolerance

Instead of the thymus, the bone marrow provides the setting for central B-cell tolerance. Pre-B cells rearrange their B-cell receptor (BCR or membrane-bound immunoglobulin) early in development. The immunoglobulin heavy (H) chain genes rearrange first, followed by light (L) chain gene rearrangement. Unproductive rearrangements and pairings leading to formation of nonfunctional immunoglobulin drive pre-B cells to apoptosis (Figure 2–6A). Functional rearrangements (functional BCRs) allow immature

surface but are resistant to Fas ligand (FasL)-mediated apoptosis because of the simultaneous expression of Bcl-xL (not shown) induced by CD28 ligation during activation. Several days after activation, when Bcl-xL presence has declined, CD4 cells become susceptible to FasL-mediated apoptosis. Natural killer cells (NK-T) may then accomplish the task of eliminating these autoreactive T cells. **2. Anergy**: Anergy may be induced via CD80/86–CD152 interaction 48 to 72 hours following activation or may result from the lack of a second costimulatory signal from APCs presenting self antigen (nonprofessional APCs). **3. Active suppression**: Active suppression is thought to occur when nonhematopoietic cells (stimulated by IFN-γ) present antigen in an MHC class II–restricted fashion to CD4 T-suppressor cells (T_S, also known as CD4 + CD25 + FOXP3 + T regs). Before becoming unresponsive, these cells may induce specific CD8 T_S cells. In turn, these CD8 T_S cells may suppress antigen-specific autoreactive T cells. **4. Ignorance (top right)**: Some autoreactive T cells may never encounter self antigen because it may be sequestered from the immune system. Although they may persist in the circulation, they never become activated. **5. Immune deviation**: Under specific circumstances, noninflammatory T_{H2} responses could suppress inflammatory (autoreactive) T_{H1} responses (see text).

A

Central B-cell tolerance

B

Peripheral B-cell tolerance
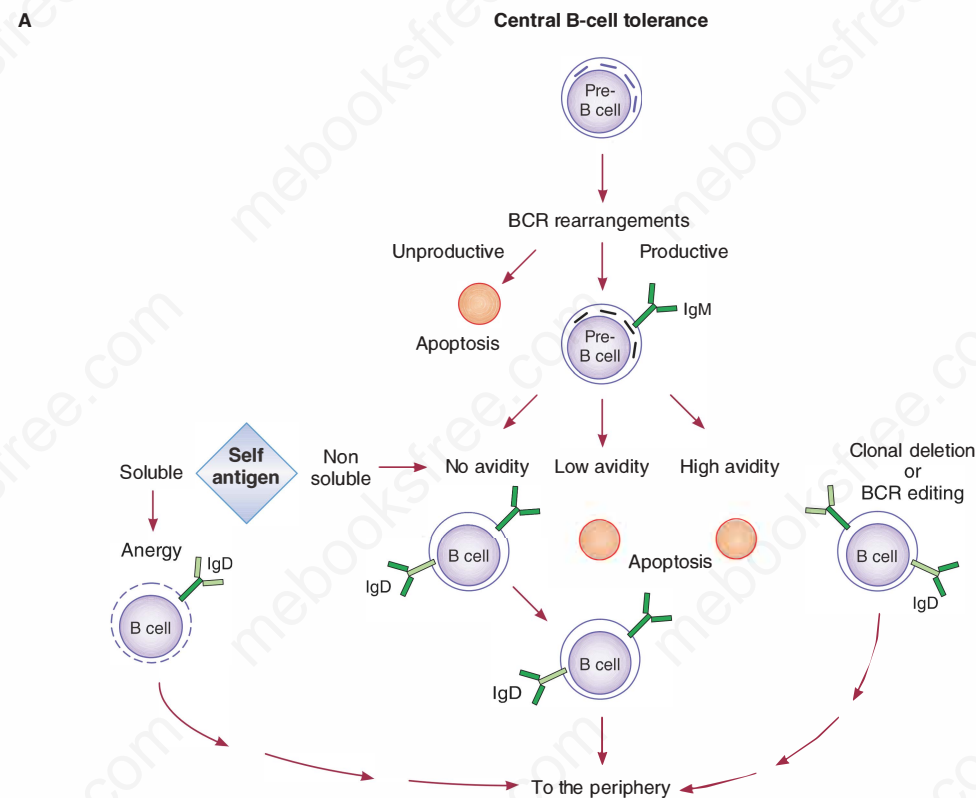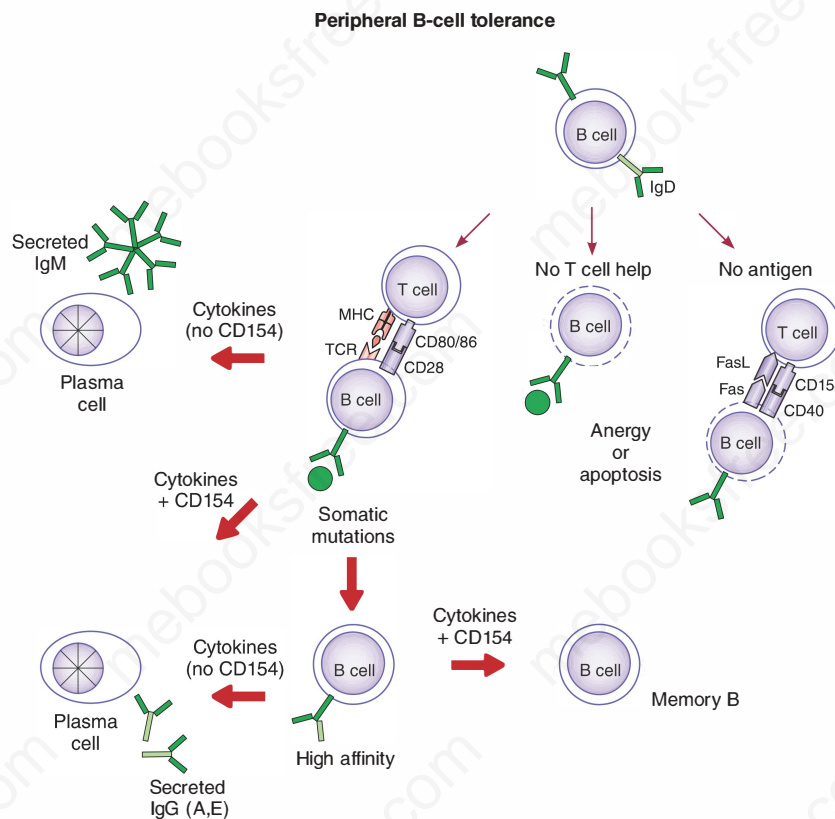
FIGURE 2–6   **A. Central B-cell tolerance**: As T cells do in the thymus, B cells rearrange their B-cell receptor (BCR) in the bone marrow. Unproductive rearrangements drive pre-B cells to apoptosis. Functional rearrangements allow expansion and expression of IgM. Next, similar to T-cell clonal deletion, immature B cells that strongly bind self antigens in the bone marrow are eliminated by apoptosis. Some autoreactive immature B cells, instead of becoming apoptotic, however, resume rearrangements of their L-chain genes, attempting to reassemble new allelic

B-cell expansion and expression of IgM and CD21 (a marker of functionality). Only one-third of the precursor cells reach this stage. The random rearrangement of the V, D, and J segments of immunoglobulin genes during this period inevitably generates self-recognizing immunoglobulins. Negative selection of autoreactive B cells occurs at the immature B cell stage on the basis of the avidity of the BCR for self antigens. Similar to the T-cell clonal deletion, immature B cells that strongly bind antigens in the bone marrow are eliminated by apoptosis. Some autoreactive immature B cells, instead of undergoing apoptosis, resume rearrangements of their L-chain genes in an attempt to reassemble new κ or λ genes. This procedure, called **BCR editing**, permanently inactivates the auto-reactive immunoglobulin genes. Soluble antigens, presumably because they generate weaker signals through the BCR of imma-ture B cells, do not cause apoptosis but render cells unresponsive to stimuli (anergy). These anergic B cells migrate to the periphery, where they express IgD. They may be activated under special cir-cumstances, making anergy less than sufficient as a mechanism of enforcing tolerance. Only immature B cells in the bone marrow with no avidity for antigens (membrane-bound or soluble) become mature B cells with the capacity to express both IgM and IgD. As with T cells, 97% of developing B cells undergo apoptosis within the bone marrow. Also, and as with T cells, **central clonal dele-tion**, **anergy**, and **BCR editing** eliminates autoreactive B cells, recognizing bone marrow-derived self antigens.

Peripheral B-cell tolerance (Figure 2–6B) is also crucial for protection against autoimmunity. It appears that in the absence of antigen, mature B cells are actively eliminated in the periphery by activated T cells via Fas–FasL and CD40–CD154 interactions. In the presence of specific antigen but without T-cell help, antigen recognition by BCRs induces apoptosis or anergy of mature B cells. If antigen and specific T-cell help are provided—that is, if antigen bound to the BCR is internalized, processed, and pre-sented in an MHC class II–restricted fashion to a previously acti-vated T_H cell specific for the same antigen—two events occur. One, the B cell becomes an IgM-secreting plasma cell, and—in the presence of the appropriate cytokines and after expression of CD40 (for T_H cell CD154 interaction)—class switching occurs. Two, further somatic hypermutation of the immunoglobulin variable region genes of such mature B cells, which changes affinity of BCRs for antigens, also occurs in germinal centers (see also Immune Recognition and Response, discussed earlier). Mutants with low-affinity receptors undergo apoptosis, while enhanced-affinity BCRs are positively selected. In the presence of CD40 ligation of CD154, antigen-stimulated B cells become memory B cells (see Figure 2–6B).

The ability of mature B cells to capture very low quantities of antigen via high-affinity BCRs allows them to amplify their anti-gen-presenting capacity to more than 1000 times that of other professional APCs. This particular property may become critical in the development of chronic organ-specific autoimmune dis-eases in which the source of antigen is limited. Thus, autoreactive B cells that happen to escape the control mechanisms described could amplify and perpetuate autoimmune responses in patients with failing endocrine organs when tissue destruction has left only minute amounts of residual antigen.

AUTOIMMUNITY IS MULTIFACTORIAL

Although the breakage of self-tolerance seems to be a central pathogenic step in the development of autoimmune diseases, autoimmunity is a multifactorial event. Specifically, defects in apoptosis-related molecules (Fas–FasL) of thymic dendritic cells have been shown to impair central clonal deletion. Also, in the periphery, similar defects (Fas–FasL, CD152) on T cell-APC mol-ecules may prevent apoptosis of autoreactive T cells. However, it is difficult to consider these general defects as causative of organ-specific disorders. Furthermore, clonal ignorance of T cells cannot be maintained if antigens sequestered from the immune system are released in blood or if cryptic epitopes of antigens that have never been recognized by the immune system are presented to T cells for recognition (after tissue destruction, for example). Defects of active suppression (T regs dysfunction, CTLA-4 downregulation), immune deviation (T_{H1}/T_{H2} imbalance), and defects in B-cell tolerance may all be involved in the pathogenesis of autoimmune diseases. How and why loss of immune self-tolerance occurs is not completely understood. Both genetic and environmental factors appear to be necessary.

GENETIC FACTORS IN AUTOIMMUNITY

Epidemiologic studies demonstrate that susceptibility to most autoimmune diseases has a significant genetic component. In type 1 diabetes, for example, there is a clear association between race and susceptibility to disease—the incidence is approximately

---

κ or λ genes (BCR editing). Soluble self antigens presumably generate weaker signals through the BCR of immature B cells; they do not cause apoptosis but make cells unresponsive to stimuli (anergy). These anergic B cells migrate to the periphery, expressing IgD, and may be activated under special circumstances. Only immature B cells with no avidity for antigens become mature B cells, expressing both IgM and IgD. These are the predominant cells that make it to the periphery. **B. Peripheral B-cell tolerance**: In the "absence" of antigen (top right), mature B cells are actively eliminated by activated T cells via Fas–FasL and CD40–CD154 interactions. In the "presence" of specific self antigen but "without T-cell help," antigen recognition by BCRs induces apoptosis or anergy on mature B cells. If self antigen and specific autoreactive T-cell help are pro-vided, two events develop (center): (1) The B cell becomes an IgM-secreting plasma cell (top left), and, in the presence of the appropriate cyto-kines after expression of CD40 (for T_H cell CD154 interaction), class switching occurs (bottom left). (2) Further somatic hypermutation of the Ig-variable region genes, which changes affinity of BCRs, occurs. Mutants with low-affinity receptors undergo apoptosis, while improved-affinity BCRs are positively selected. In the presence of CD40 ligation of CD154, antigen-stimulated B cells become memory B cells. These two events are the same as in foreign antigen recognition.

40 times higher in Finland than in Japan. Family studies also demonstrate a strong underlying genetic component. The lifetime risk of developing type 1 diabetes in the United States general population is 0.4%, whereas in relatives of diabetics the risk is substantially higher (4% for parents, 5%-7% for siblings, 20% for HLA-identical siblings, 25%-40% for monozygotic twins).

The inheritance pattern of autoimmune disorders is complex. These disorders are polygenic, arising from several independently segregating genes. The most consistent genetic marker for autoimmune diseases to date is the MHC genotype. Considering again genetic susceptibility to type 1 diabetes, up to 95% of Caucasians developing diabetes express the HLA alleles DR3 or DR4—compared to about 40% of normal individuals. Individuals heterozygous for both DR3 and DR4 have the highest risk. It has been shown that the DQ rather than DR genotype is a more specific marker of susceptibility and that the association of both markers is due to the fact that they are products of closely linked genes. But what is more important than the fact that HLA genes are linked to diabetes is that HLA haplotypes are no longer simply undefined genetic markers. It has been shown that the polymorphisms of the DQ molecules are critical for high-affinity recognition of autoantigens (eg, islet cell antigens) by TCRs. HLA-DQ structure analysis suggests that the lack of aspartic acid at position 57 (Asp^{57}) on the DQ β chain allows the autoantigen (processed peptide) to fit better in the antigen-binding groove formed by this molecule. On the contrary, the presence of Asp^{57} allows the formation of a salt bridge with a conserved arginine at position 76 on the DQ α chain, preventing the accommodation of the immunogenic peptide recognized by the TCR. Several autoimmune diseases have been linked to HLA-DQβ1 genes, including type 1 diabetes, celiac disease, bullous pemphigoid, autoimmune hepatitis, and premature ovarian failure, and the structure of the DQβ1 molecule may be the reason for the increased susceptibility. Other candidate genes associated with autoimmune endocrinopathies are discussed further under the single and polyglandular syndromes.

ENVIRONMENTAL FACTORS IN AUTOIMMUNITY

Environmental factors also play a critical role in the pathogenesis of autoimmune disease. The strongest evidence for this statement comes from studies of monozygotic twins, which show that concordance rates for autoimmune disorders are imperfect (never 100%). As mentioned earlier, in type 1 diabetes, identical twins show less than 50% concordance.

The environmental factors thought to have greatest influence on disease development include infectious agents, diet, and toxins. In type 1 diabetes, viruses have been strong suspects. Up to 20% of children prenatally infected with rubella develop type 1 diabetes. Children with congenital rubella also have an increased incidence of other autoimmune disorders, including thyroiditis and dysgammaglobulinemia. The mechanisms by which these pathogens may induce autoimmune responses include molecular mimicry and direct tissue injury. The hypothesis of molecular mimicry suggests that immune responses directed at infectious agents can cross-react with self antigens, causing tissue or organ destruction. Support for this concept is found in well-known clinical syndromes such as rheumatic fever (immune responses directed against streptococcal M protein seem to cross-react with cardiac myosin, inducing clinical myocarditis). In autoimmune diabetes, the best-studied example of molecular mimicry is the B4 coxsackievirus protein P2-C. Coxsackie B4 virus has also been epidemiologically implicated in the development of type 1 diabetes. There is a striking amino acid sequence similarity between P2-C viral protein and the enzyme glutamic acid decarboxylase (GAD), found in pancreatic β cells (see Autoimmune Aspects of Type 1 Diabetes, discussed later).

The importance of diet in the development of autoimmune diseases remains controversial. An association between early exposure to cow's milk proteins and the risk of type 1 diabetes has been observed in several epidemiologic studies. For example, one study demonstrated that primary immunity to insulin is induced in infancy by oral exposure to cow's milk insulin, but the relevance of this observation is still unknown. On the other hand, selected antigens (from bovine serum albumin to porcine insulin) have been administered orally to mice with a broad spectrum of autoimmune disorders, including nonobese diabetic (NOD) mice, with favorable outcomes. Those data in mice were so compelling that oral tolerance trials in humans have been conducted or are ongoing. Unfortunately, the results of already completed trials in other autoimmune diseases have been disappointing. Three large, randomized, controlled trials designed to delay or prevent type 1 diabetes—the two Diabetes Prevention Trial (DPT-1 and -2) and the European Diabetes Nicotinamide Intervention Trial—have failed to demonstrate a treatment effect. Thus, it should not be concluded that it is impossible to delay or prevent type 1 diabetes; rather, it may require testing of more potent interventions or combinations of therapies, guided by better understanding of the immunopathogenesis of the disease, to demonstrate attenuation or amelioration of the destructive immune process leading to type 1 diabetes.

SINGLE-GLAND AUTOIMMUNE SYNDROMES

Organ-specific autoimmune endocrine disorders may present as single entities or may cluster in polyendocrine syndromes. Most endocrine glands are susceptible to autoimmune attack. Some are affected more frequently than others (see Table 2–1).

AUTOIMMUNE ASPECTS OF THYROID DISEASE

Autoimmune thyroid disease can present in a polarized fashion with Graves disease (thyroid hyperfunction) at one end and Hashimoto thyroiditis (thyroid failure) at the other. This functional subdivision is clinically useful. However, both diseases have a common autoimmune origin.

Genes and Environment

Major susceptibility genes in autoimmune thyroid disease have yet to be identified. Although certain HLA alleles (mainly HLA-DR3 and DQA1*0501) have been shown to be present more frequently in Graves disease than in the general population, this association has frequently been challenged. In fact, no consistent association has been found between Graves disease and any known HLA polymorphism. Furthermore, the risk of developing Graves disease in HLA-identical siblings (7%) is not significantly different from that in control populations. HLA-DR5, -DR3, -DQw7 in Caucasian, HLA-DRw53 in Japanese, and HLA-DR9 in Chinese patients were found to be associated with Hashimoto thyroiditis. However, genetic linkage between Hashimoto thyroiditis and a specific HLA locus has not been demonstrated consistently either. Overall, the HLA loci are likely to provide less than 5% of the genetic contribution to autoimmune thyroid disease, confirming the relative importance of non-HLA-related genes in susceptibility. For example, it has been shown that the inheritance pattern of autoantibodies to thyroperoxidase (TPO) is genetically transmitted. Other candidates are currently under study. However, autoimmune thyroid disease linkage to CTLA-4, HLA, IgH chain, TCR, thyroglobulin (Tg), TPO, and thyrotropin receptor (TSH-R) genes has been excluded. Using positional genetics of a candidate gene, Graves disease susceptibility was mapped to a noncoding 6.13 kb 3′ region of CTLA-4, the common allelic variation of which was correlated with lower messenger RNA levels of the soluble alternatively spliced form of CTLA-4 (sCTLA-4). sCTLA-4 is known to be present in human serum. It can bind CD80/86 molecules on APCs and inhibit T-cell proliferation *in vitro*. The reduction in the level of sCTLA-4 could potentially lead to reduced blocking of CD80/86, causing increased activation through CD28 of autoreactive T cells.

An important environmental factor influencing the natural history of autoimmune thyroid disease is that of iodine intake (dietary, or present in drugs such as amiodarone or in x-ray contrast media). There is considerable evidence that iodine adversely affects both thyroid function and antibody production in those with occult or overt autoimmune thyroid disease.

Autoimmune Response

In Graves disease, thyrocytes are the differentiated carriers of TSH-Rs and the target cells of autoantibodies and probably the autoimmune response. The development of autoantibodies that functionally stimulate the TSH-R mimicking the action of TSH was the first example of antibody-mediated activities of a hormone receptor in humans. Autoantibodies that may stimulate the calcium-sensing receptor (CaSR) (another G-protein–coupled receptor) and signal the inhibition of PTH production have been described in autoimmune hypoparathyroidism. Similarly, stimulating antibodies that bind to the adrenocorticotropin (ACTH) receptor may be involved in the pathogenesis of primary pigmented nodular adrenocortical disease (also referred to as nodular adrenal dysplasia).

In Graves disease, antibodies to the TSH-R present with different types of activity. Thyroid-stimulating immunoglobulins (TSIs), the cause of the hyperthyroidism, are detected by a bioassay that measures cAMP production in a cell line that expresses TSH-Rs. TSH-R autoantibodies (stimulating, blocking, or neutral) can be identified by their ability to prevent TSH binding to the TSH-R (TSH-binding inhibition immunoglobulins [TBII]). No direct immunoassay for the measurement of TSH-R autoantibodies is available yet, and its development may be difficult because TSH-R autoantibodies are present at very low concentrations in patients with the disease.

A particular feature of Graves disease is its early clinical presentation. Unlike other autoimmune endocrinopathies (type 1 diabetes, Hashimoto thyroiditis, autoimmune Addison disease), in which much of the target organ has to be destroyed before the disease is manifested, Graves hyperthyroidism often presents with an enlarged and active gland. Minimal lymphocytic infiltration is present when hyperthyroidism (due to the presence of TSH-R-stimulating antibodies) develops. This unique feature may ultimately allow early immune intervention in preference to current ablative therapeutic options.

Another peculiar feature of Graves disease is the helper T-cell response observed in this disease. The activation of antibody-producing B cells by T_H lymphocytes in Graves disease is well recognized. At present, a prevailing concept of human autoimmunity suggests that, as in acute allograft rejection, "deviation" toward a T_{H1}/T_{H17} response dominates its pathogenesis. Counter-deviation toward T_{H2} is thought to be a consequence of tolerance induction and has been postulated as a potential therapeutic approach. Graves disease seems to challenge that concept. Analysis of TSH-R-specific T-cell clones from patients with Graves disease has provided direct evidence for polarization of T_H responses; however, instead of T_{H1} deviation, T_{H0} and T_{H2} responses have been observed. As mentioned before, T_{H1} cells, which regulate cell-mediated responses, secrete mainly IFN-γ and small amounts of IL-4. In contrast, T_{H2} cells that regulate antibody production (such as TSH-R autoantibodies in Graves disease) preferentially produce IL-4 and little IFN-γ. T cells expressing both IL-4 and IFN-γ are known as T_{H0} cells. These experimental results suggest that in Graves disease T_{H0} to T_{H2} cell responses appear to be dominant. Hence, in human autoimmunity, Graves disease appears to be an exception to the usual T_H cell pattern.

In Hashimoto thyroiditis, the hallmark of the humoral immune response is the presence of autoantibodies to TPO. Although the effector mechanism for TPO (or thyroglobulin [Tg]) autoantibodies is still controversial, under special circumstances (at least *in vitro*) the autoantibodies are themselves cytotoxic agents or activators of cytotoxic T lymphocytes. Furthermore, in secondary T-cell responses, antibodies may play a critical role in antigen processing or presentation to T cells. In short, macrophages internalize (and subsequently process) antigen by phagocytosis and antigen-antibody complex uptake via Fc receptors. B cells have membrane-bound antibodies (B-cell receptors [BCRs]) which provide a much more powerful system for antigen capture. Indeed, recombinant TPO-specific membrane-bound autoantibody captures antigen and allows presentation efficiently. Antibody binding also modulates antigen processing of immune complexes, enhancing or suppressing the presentation of different

T-cell peptides. Hence, APCs (internalizing immune complexes through Fc receptors) and B cells (capturing antigen through BCRs) can influence the secondary T-cell response that perpetuates autoimmune disease. The potential role of autoantibodies in modulating presentation of T-cell determinants in thyroid (and diabetes) autoimmunity is being explored.

Animal Models of Autoimmune Thyroid Disease

The classic immunologic approach to development of an animal model of an autoimmune disease is to immunize the animal with soluble antigen in adjuvant. For autoimmune thyroid disease, the induction of thyroiditis in rabbits using human Tg was one of the earliest attempts to do this—by Rose and coworkers in 1956. In subsequent studies, mice immunized with human or murine Tg developed experimental autoimmune thyroiditis. Immunization with TPO (human or porcine) induces thyroid autoantibodies and, as in the case of Tg, causes thyroiditis to develop in particular MHC strains of mice. However, unlike spontaneous thyroiditis in chickens, none of the immunized mouse models of thyroiditis develop hypothyroidism.

In 1996, Shimojo and coworkers developed a mouse model that clearly mimics some of the major features of Graves disease. This was achieved by the ingenious approach of immunizing mice with fibroblasts stably transfected with the cDNA for the human TSH-R and syngeneic MHC class II (Figure 2–7). Most of the animals had moderately high TBII activity in their sera, and about 25% were clearly thyrotoxic, with elevated T_4 and T_3 values, detectable TSI activity, and thyroid hypertrophy. For the first

time, therefore, an animal model was established in which a significant number of affected subjects have the immunologic and endocrinologic features of Graves hyperthyroidism. More recently, other models have become available. The general rule for induction of an antibody response to the TSH-R in the form of TSI and consequent Graves disease appears to be the need to express the antigen in a native form in an MHC class II carrying cell. Thus, TSH-R-transfected B cells (M12 cell line), TSH-R-adenovirus-infected dendritic cells, and even naked TSH-R DNA or TSH-R adenovirus parenterally delivered and probably captured and expressed by MHC class II expressing cells, have been used to induce TSI and Graves-like disease. However, the focal thyroiditis that accompanies human Graves, as well as the extrathyroidal manifestations that define the disease, have never been reliably reproduced in any of these models. Nevertheless, these models open up new ways of investigating the pathogenesis of Graves disease.

AUTOIMMUNE ASPECTS OF TYPE 1 DIABETES

Most type 1 diabetes results from autoimmune destruction of pancreatic β cells in a process that can span several years. This results in glucose intolerance and clinical disease when the majority of β cells have been destroyed. The destruction is marked by circulating antibodies to pancreatic β cells and by massive infiltration of mononuclear lymphocytes into the islets of Langerhans while pancreatic β cells remain. The lymphocytes slowly disappear when the β cells are gone. Although insulin is available for replacement therapy, type 1 diabetes remains a chronic disorder of major socioeconomic impact, especially because it mainly afflicts the young. Elucidation of the molecular mechanisms underlying this destruction—and the development of methods to prevent autoimmunity—may ultimately lead to effective treatment. Such developments, however, require animal models of type 1 diabetes that closely resemble the disease in humans.

Genes and Environment

The susceptibility to develop type 1 diabetes is associated with certain alleles of the MHC class II locus that have been statistically linked to a variety of autoimmune disorders. The most recent analyses indicate that in Caucasians, HLA-DR3, DQ2 (DQB1*0201) and HLA-DR4 (DRB1*0401), DQ8 (DQB1*0302) haplotypes are most strongly associated with type 1 diabetes. In Asian populations, DRB1*0405 is the major susceptibility haplotype. In contrast, the DR2, DQ6 (DQB1*0602) haplotype is negatively associated with type 1 diabetes. More importantly, susceptibility requires both HLA-DQ β chain alleles to be negative for Asp[57] in the amino acid sequence. Studies of different populations have shown a linear relationship between the incidence of type 1 diabetes and the estimated frequency of homozygous absence of Asp[57].

Non-HLA candidate genes consistently associated with type 1 diabetes include the "variable number of tandem repeats" (VNTR)

hTSH-R + m class II fibroblasts

TSH-R

MHC

TSH-R peptide

Injected cells

AKR/N mice

TSH-R antibodies and hyperthyroidism

FIGURE 2–7 Animal model for Graves disease. This first animal model of Graves-like disease was achieved by injecting AKR/N mice with syngeneic (MHC-identical) fibroblasts dually transfected with mouse MHC class II (H2-k) and human TSH-R cDNA. About 25% of the animals developed endocrinologic (hyperthyroidism) and immunologic (TSI, TBII) features of Graves disease.

polymorphisms in the insulin gene and the CTLA-4 gene (CD152). The VNTR polymorphisms are located adjacent to defined regulatory sequences that influence insulin gene expression. Of immunologic importance, CTLA-4 gene (see Immune Recognition and Response, discussed earlier) is the other non-HLA candidate gene consistently found to be associated with type 1 diabetes.

Although environmental factors definitely play a role in the development of type 1 diabetes (eg, Coxsackie B4 virus, mumps virus, rubella virus, Kilham rat virus in the bio breeding [BB]-rat; or cow's milk formula exposure), more studies are needed to establish a definite etiologic link (see also Genetic and Environmental Factors in Autoimmunity earlier).

Autoimmune Response

The autoantibodies associated with β cell destruction can be present up to several years before the clinical onset of disease and are thus excellent markers of disease risk. Furthermore, they have served as important tools to identify human pancreatic β cell autoantigens. In 1990, Baekkeskov and coworkers identified a 64-kDa islet cell protein as the smaller isoform of the enzyme that synthesizes γ-aminobutyrate (GABA): glutamic acid decarboxylase (GAD65). This autoantigen was shown to be recognized by 70% to 80% of prediabetic and newly diagnosed type 1 diabetic patients' sera. A second component of the 64-kDa antigen was shown to be a putative tyrosine phosphatase, termed IA-2. IA-2 is recognized by 60% to 70% of prediabetic and newly diagnosed type 1 diabetic patients. Together, GAD65 and IA-2 autoantibodies detect over 90% of individuals who develop type 1 diabetes and can be used to detect individuals at risk several years before the clinical onset of disease.

Although autoantibody responses to GAD65 are not easily detected, there is strong evidence to suggest that GAD65 is an important T-cell autoantigen in the NOD mouse. Thus, GAD65 is the earliest known target of the autoimmune T-cell response in the NOD mouse. Administration of the protein in a tolerogenic form prevents disease in NOD mice. In contrast, induction of tolerance to other potential autoantigens in this model (such as carboxypeptidase H and hsp60) does not prevent disease. The NOD mouse does not develop autoimmunity to the IA-2 molecule and thus distinguishes itself from the human disease with regard to this target antigen (see Models, discussed later).

Insulin is a third well-characterized autoantigen in type 1 diabetes. Insulin autoantibodies can be detected in about 50% of newly diagnosed children with type 1 diabetes. Insulin-specific T-cell clones can transfer disease in the NOD mouse. Furthermore, administration of whole insulin, insulin B chain, or an insulin peptide epitope in a tolerogenic form can protect against disease in NOD mice. Because animals receiving insulin or insulin B chain continue to have intra-islet insulitis—in contrast to young NOD mice treated with GAD65 in a tolerogenic way—it has been suggested that insulin reactivity is more distal in disease progression. Additional but less well-characterized proteins have been implicated as targets of autoantibodies in type 1 diabetes in humans.

Autoantibodies, although they are good markers of disease, do not seem to be directly involved in destruction of pancreatic β cells. Adoptive transfer of diabetes to NOD mice with spontaneous, combined immunodeficiency (NOD-SCID) lacking B cells, can be mediated by T cells alone. However, because β-cell-deficient NOD mice do not develop disease, it is possible that B lymphocytes function as important APCs in the islet to perpetuate an ongoing autoimmune response and thus are essential for presentation of rare antigens such as GAD65 and IA-2. (See also autoimmune response in the section on Autoimmune Aspects of Thyroid Disease, discussed earlier.)

An important question is whether GAD65, IA-2, and insulin are major target antigens of T-cell mediated β-cell destruction that results in type 1 diabetes in humans. Proliferative and cytotoxic T-cell responses to GAD65 are detected in the peripheral blood of newly diagnosed type 1 diabetes patients, but their pathogenicity has not been addressed. Induction of neonatal tolerance to GAD65 specifically prevents diabetes in the NOD mouse model. The role of IA-2 in destructive autoimmunity to the pancreatic β cell in humans is suggested by the high-predictive value of IA-2 antibodies for clinical onset of diabetes.

Both GAD65 and IA-2 are neuroendocrine proteins, which are expressed at significant levels in the brain and β cell. Stiff-man syndrome—a very rare neurologic disorder in humans with a high coincidence of diabetes—is characterized by a strong autoantibody response to GAD65, the titer of which is several orders of magnitude higher than in diabetes. It has been suggested that impairment of GABA-secreting neurons in stiff-man syndrome is mediated by GAD65 autoantibodies, whereas development of type 1 diabetes is associated with a cellular immune response to GAD65. The low incidence of stiff-man syndrome compared with type 1 diabetes (only one in 10^4 type 1 diabetic patients develops stiff-man syndrome, whereas 40% of stiff-man syndrome patients develop type 1 diabetes) probably reflects in part the protection of GABAergic neurons by the blood–brain barrier and the absence of MHC class II antigen expression in normal neurons. The cellular localization of IA-2 expression in brain is not known, and there are no known disorders of the central nervous system that involve autoimmunity to IA-2.

In the NOD mouse, the destruction of pancreatic β cells requires both the CD4 T-helper (T_H) cells and CD8 cytotoxic (T_C) cells. Whereas T_H cells seem to be required for the development of an autoimmune response to the islets and generation of intrainsulitis, T_C cells are probably the effector cells of β-cell destruction. Furthermore, there is evidence that in the CD4 lineage, the T_{H1} subset is important for development of disease in the NOD mouse. T_{H1} cells are induced by IL-12 and are biased toward secreting IFN-γ and IL-2. In contrast, there is evidence that the T_{H2} cytokine IL-4 exerts a dominant-negative effect on diabetes progression in the NOD mouse. In humans, low autoantibody titers associated with type 1 diabetes and high titers associated with a protective haplotype (DR2) suggest that a strong T_{H2} response can be inhibitory for β-cell destruction. A role for T_{H1} cells in human disease is also suggested by results of cytokine profiles of peripheral human NK cells in identical twins which are discordant for the development of diabetes. This is different from

the observed T_H responses in Graves disease (see Autoimmune Aspects of Thyroid Disease discussed earlier).

It has been suggested that it is not the presence of autoantibodies to GAD65 but the absence of the corresponding anti-idiotypic antibodies that defines type 1 diabetes. Anti-idiotypic antibodies bind to the idiotype (binding region) of other antibodies. Investigators found that while diabetes patients are positive for GAD65 antibodies because they lack certain anti-idiotypic antibodies, healthy individuals are negative for GAD65 antibodies because of the very presence of anti-idiotypic antibodies directed against GAD65 antibodies in their serum.

Animal Models of Autoimmune Diabetes Mellitus

The NOD mouse has been invaluable for studies of molecular mechanisms of autoimmunity directed toward the pancreatic β cells and the development of diabetes. It has several features, however, which distinguish it from the human disease. The incidence of diabetes is two to three times higher in female than in male NODs, whereas in humans there is a slight preponderance of type 1 diabetes in males. Furthermore, while the induction of organ-specific autoimmunity and inflammation in humans may be caused by human pathogens or toxins, autoimmunity seems to be the default mechanism in the NOD mouse. Thus, mice in clean, pathogen-free environments have a high incidence of disease, whereas a variety of regimens that stimulate the immune system of the mouse, such as viral infection or injection of complete Freund adjuvant, prevent disease. To date, more than 125 treatments for successful prevention or delay of diabetes in the NOD mouse have been identified, but none have been identified for humans.

The BB rat develops spontaneous T-cell-mediated diabetes. The BB-rat disease is significantly different from the human disease in that it is accompanied by autoantibodies to lymphocytes and a severe lymphocytopenia, which is essential for development of β-cell autoimmunity and diabetes in this model.

In an attempt to develop better models of diabetes, "humanized" transgenic mice that express diabetes-prone **human MHC class II** molecules were developed. Since these animals did not develop spontaneous diabetes, they were backcrossed into the NOD background. However, backcrossing to NOD once again prevented the development of diabetes. Other animal models of type 1 diabetes, some of them *carrying* **human MHC class II** diabetes-susceptibility genes, were developed by expression of ectopic molecules in the islets using the rat insulin promoter (RIP). In some of these models, diabetes needed to be induced by immunization with, for example, insulin peptides or lymphocytic choriomeningitis virus. However, none of these models fully reflect the human disease in that the immune attack was initiated against target human β-cell autoantigen(s) in an MHC class II–restricted fashion, without the need for *aberrant protein(s)* (foreign to the pancreatic β-cell) expression.

Although some of the models of type 1 diabetes described have been very useful for studies of basic immunologic mechanisms associated with pancreatic β-cell autoimmunity, other models of type 1 diabetes closer to what occurs in humans are needed if immunoprevention and immunomodulatory techniques are to be tested. At a minimum, human susceptibility genes and human target antigens need to coexist in an animal model to mimic human autoimmune responses.

AUTOIMMUNE ASPECTS OF OTHER ENDOCRINOPATHIES

Autoimmune Adrenal Failure

Autoimmune Addison disease seldom develops as a single-gland syndrome. In about 50% of cases, the disease is associated with other gland and organ failure. Anderson and coworkers described the existence of adrenal-specific autoantibodies for the first time in 1963. Using immunofluorescence techniques on sections of human, bovine, or monkey adrenals, antibodies specifically recognizing the adrenal cortex (ACA) were described. Steroid-producing cell autoantibodies (SCA) reactive with cells of the adrenals, gonads, and placenta were described by Anderson and coworkers in 1968. SCAs are detected predominantly in ACA-positive patients with Addison disease who have premature ovarian failure in the context of autoimmune polyendocrine syndrome type I (APS-I; see Autoimmune Polyendocrine Syndromes, discussed later).

Steroid 21-hydroxylase has been identified as a major adrenal autoantigen in ACA-positive patients with Addison disease. Using a sensitive assay based on the immune precipitation of radiolabeled recombinant 21-hydroxylase, workers in one study reported positive testing in sera from 72% of patients with isolated Addison disease, 92% of patients with APS-1, 100% of patients with APS-2, and 80% of patients who were positive for ACA by immunofluorescence but did not have clinically overt Addison disease (apparently healthy blood donors showed 2.5% positivity) (Table 2–2). Another study measured ACA in 808 children with organ-specific autoimmune diseases without adrenal insufficiency. ACAs were detectable in 14. Ten of these ACA-positive children (also positive for 21-hydroxylase antibodies) and 12 ACA-negative children were prospectively followed with adrenocortical function testing and antibodies. Overt Addison disease developed in 9 (90%) ACA/21-hydroxylase antibody-positive children within 3 to 121 months, and the one remaining

TABLE 2–2 Adrenal autoantibodies in different syndromes.

Autoantibodies	Addison Disease (%)	APS-1[a] (%)	APS-2[b] (%)	ACA(+)[c] (%)
21-Hydroxylase	72	92	100	80
17α-Hydroxylase	5	55	33	20
P450scc	9	45	36	20

[a]APS-1: Autoimmune polyglandular syndrome type 1: Autoimmune polyendocrinopathy, candidiasis, and ectodermal dystrophy.

[b]APS-2: Autoimmune polyglandular syndrome type 2: Adrenal insufficiency, thyroid disease, and diabetes mellitus.

[c]ACA(+): Adrenal cortex antibody-positive without clinically overt Addison disease.

child had subclinical hypoadrenalism throughout an extra 24-month observation period. The progression to adrenal failure was not related to ACA titer, sex, baseline adrenal function (subclinical insufficiency vs normal function), type of associated autoimmune disease, or HLA status. Although ACA 21-hydroxylase antibodies appear to be highly predictive in children, in adults the cumulative risk of developing Addison disease in patients with other organ-specific autoimmune diseases and positive ACA 21-hydroxylase antibodies is about 32%. Steroid 17α-hydroxylase is another adrenal autoantigen. 17α-hydroxylase antibodies were found in 5% of patients with isolated Addison disease, 55% of patients with APS-1, 33% of patients with APS-2, and 20% of sera from patients who were positive for ACA but did not have clinically overt Addison disease (see Table 2–2). Antibodies against another adrenal autoantigen, cytochrome P450 side-chain cleavage enzyme (P450scc), were found to be present in 9% of patients with isolated Addison disease, 45% of patients with APS-1, 36% of patients with APS-2, and 20% of sera from patients who were positive for ACA but did not have clinically overt Addison disease (see Table 2–2). The prevalence of P450scc antibodies in these groups of patients was always lower than that of 21-hydroxylase antibodies but similar to that of 17α-hydroxylase antibodies. Furthermore, almost all sera that were positive for 17α-hydroxylase or P450scc antibodies were also positive for 21-hydroxylase antibodies. In addition, a comparison of SCAs measured by immunofluorescence with 17α-hydroxylase and P450scc antibody measurements suggested that 17α-hydroxylase and P450scc are the major components of the SCA antigen just as 21-hydroxylase is the major component of ACA antigen.

Overall, immune responses in autoimmune adrenal disease may involve other antigens, but reactivity to the three described, particularly 21-hydroxylase, appears to predominate. Although inhibition of enzymatic activity by these antibodies has been shown *in vitro*, no clear relationship to the pathogenesis of the clinical syndrome has yet been established.

Autoimmune Oophoritis and Orchitis

An autoimmune origin for premature ovarian failure with concomitant Addison disease or oophoritis can be based on the following human and animal data: (1) the presence of autoantibodies to SCA in most cases, (2) the characterization of shared autoantigens between the adrenals and the ovaries (ie, 17α-hydroxylase and P450scc), (3) the histologic features of the ovaries (lymphocyte and plasma cell infiltrate involving steroid-producing cells), and (4) animal models of the syndrome. There is some evidence of autoimmunity in idiopathic premature ovarian failure not associated with Addison disease (cellular immune abnormalities, presence of various ovarian antibodies in some patients, and associations with type 1 diabetes and myasthenia gravis); however, the absence of histologic confirmation makes the autoimmune pathogenesis less credible.

Less is known about the autoimmune pathogenesis of human orchitis. Animal models, however, have shown that infectious or traumatic injury to the testes can induce autoimmune responses in this immune-privileged tissue (defined as sites where antigens are not presented to the host immune system; see Tolerance, discussed earlier).

Autoimmune Hypophysitis

Autoimmune hypophysitis (also called lymphocytic hypophysitis) should be considered in the differential diagnosis of pituitary abnormalities in women (8:1 female:male ratio) during the latter half of pregnancy and in the first 6 months postpartum, as well as in patients with coexisting autoimmune disorders, for example, thyroiditis, adrenalitis, autoimmune hypoparathyroidism, or atrophic gastritis. More than 380 cases have been described since the original report in 1962. Antipituitary antibodies have been detected in a minority of patients. Owing to the lack of markers for the disease, the diagnosis can only be confirmed with histologic examination. Nevertheless, because of the usually transient endocrine and compressive features of this condition, conservative management based on clinical suspicion may prevent the consequences of unnecessary pituitary surgery. Granulomatous hypophysitis—another form of autoimmune hypophysitis—appears to have a similar autoimmune pathogenesis but more commonly affects postmenopausal women and men. The presence of T regs in this form of hypophysitis, however, makes the autoimmune pathogenesis less clear.

Autoimmune Hypoparathyroidism

Autoimmune hypoparathyroidism, also called idiopathic hypoparathyroidism, is one of the major components of APS type 1 (APS-1; see next section). It also presents as a sporadic disease, sometimes associated with Hashimoto thyroiditis in women. The fact that autoimmune hypoparathyroidism presents in association with other autoimmune diseases and also the presence of autoantibodies reactive with parathyroid tissue in many affected patients suggests an autoimmune pathogenesis. Parathyroid autoantibodies have been reported to show a complement-dependent cytotoxic effect on cultured bovine parathyroid cells. At least one major parathyroid autoantigen has been identified as the CaSR. The CaSR is of great importance in the regulation of parathyroid hormone secretion and renal tubular calcium reabsorption. This receptor is a member of the 7-membrane-spanning domain G protein–coupled receptor family. It is also expressed in thyroid C cells, the pituitary, the hypothalamus, and in other regions of the brain. The relationship of the autoimmune response directed against the receptor to the pathogenesis of the disease is not clear. However, antibody-mediated stimulation of the CaSR with consequent inhibition of PTH synthesis and secretion has been suggested. The prevalence of the CaSR antibodies in clinically diagnosed idiopathic hypoparathyroidism was found to be 56% in one study. Measurement of these antibodies may have value in predicting the development of autoimmune hypoparathyroidism in patients with autoimmune endocrinopathies who are at risk. Furthermore, hypercalcemia in a patient with multiple autoimmune disorders, responsive to glucocorticoids, has been described as secondary to the presence of a blocking IgG4 autoantibody

directed against the CaSR and apparently capable of blocking the inhibitory actions of the ligand calcium (see also Chapter 8).

AUTOIMMUNE POLYENDOCRINE SYNDROMES

Associations of multiple autoimmune endocrine disorders have been classified in different syndromes. APS type 1 and type 2 (APS-1 and -2) can be clearly separated clinically (Table 2–3). Some authors have attempted to subdivide APS-2 (ie, APS-2 and -3) on the basis of the association of some autoimmune disorders but not others. Little information is gained, however, by making this subdivision in terms of understanding pathogenesis or prevention of future endocrine failure in patients or their relatives. Other autoimmune associations, not always described in syndromes, are also classically recognized. Vitiligo, for example, seems to

accompany multiple autoimmune endocrinopathies. There is now convincing evidence of linkage between *NALP1* (*NACHT leucine-rich-repeat protein 1*), a gene involved in immune regulation, and the presence of vitiligo associated with at least one endocrine autoimmune disease including but not limited to type 1 diabetes, Addison disease, and thyroiditis.

AUTOIMMUNE POLYENDOCRINE SYNDROME 1 (APS-1)

APS-1 is an autosomal recessive disorder with 25% incidence among siblings of affected individuals. Also known as APECED, or autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy, APS-1 is characterized by the triad of chronic mucocutaneous candidiasis, autoimmune hypoparathyroidism, and adrenal insufficiency (only two are required in the index case for the

TABLE 2–3 Comparison of the different components of autoimmune polyendocrine syndromes.

Characteristics	Type 1	Type 2
Inheritance	Autosomal recessive	Polygenic
Genetic association or linkage	Linked to AIRE	Some HLA association
Gender	Equal distribution	Female preponderance
Age at onset	Infancy	Age 20-40
Endocrine disorders		
Addison disease	60%-72%	70%
Hypoparathyroidism	Common (79%-96%)	Rare (late onset)
Autoimmune thyroid disease	Less frequent (about 5%)	More frequent (about 70%)
Type 1 diabetes	14% (lifetime)	>50%
Primary hypogonadism	60% female, 14% male	About 5%
Hypophysitis	Not reported	Reported
Dermatologic		
Chronic mucocutaneous candidiasis	Often at onset (about 100%)	Not reported
Alopecia	Common (about 29%)	Reported
Vitiligo	About 13%	About 5%
Dermatitis herpetiformis	Not reported	Reported
Gastrointestinal		
Celiac disease	None (only steatorrhea)	Present in 2%-3%
Autoimmune hepatitis	About 12%	Not reported
Hematologic		
Pernicious anemia	About 13%	As common as in APS-1
Pure red cell hypoplasia	Reported	Not reported
Idiopathic thrombocytopenic purpura	Not reported	Reported
Ectodermal		
Enamel hypoplasia	All reported	Not reported
Nail dystrophy		
Tympanic membrane calcification		
Neurologic		
Myasthenia gravis	None reported	All reported
Stiff-man syndrome		
Parkinson disease		
Other		
Asplenism	Reported	Not reported
Keratopathy	Reported	Not reported
Progressive myopathy	Reported	Not reported
IgA deficiency	Not reported	Reported
Serositis	Not reported	Reported
Idiopathic heart block	Not reported	Reported
Goodpasture syndrome	Not reported	Reported

diagnosis, and only one in the siblings). Chronic mucocutaneous candidiasis (involving oral mucosa and nails or, less frequently, the esophagus) is usually manifested as the initial problem early in life. In most individuals, the development of autoimmune hypoparathyroidism, a major clinical phenotype, usually follows. Specifically for hypoparathyroidism, the presence of antibodies to NALP5 (NACHT leucine-rich-repeat protein 5), a gene highly expressed in parathyroid and ovary, have been described. In one study these antibodies were detected in 49% of patients with known APS-I and hypothyroidism (see also Chapter 8). Addison disease is another component of the triad that can manifest prior to, concomitantly with, or following hypoparathyroidism. Lifelong surveillance is important since decades may elapse between the development of one feature of the disorder and the onset of another. There is no female preponderance in this syndrome, and it is not HLA-associated. APS-1 may occur sporadically or in families. The genetic locus responsible for the disease has been mapped to the long arm of chromosome 21 (21q22.3). The haplotype analysis of this region in different populations has shown that APS-I is linked to different mutations in a gene identified as the autoimmune regulator (AIRE). AIRE encodes a putative nuclear protein with transcription factor motifs (including two zinc finger motifs). It is expressed in different tissues but particularly in the thymus. The mechanism by which mutations of this putative transcription factor lead to the diverse manifestations of APS-1 is still being defined. In mice, however, the absence of the analogous protein "aire" influences ectopic expression of peripheral tissue antigens in thymic medullary epithelial cells (MECs), resulting in the development of an autoimmune disorder similar to APS-1 and establishing aire/AIRE as an important factor in the induction of central tolerance. Aire has been reported to promote the perinatal generation of a distinct compartment of Foxp3+CD4+ T reg cells, which stably persists in adult mice. This population has a role in maintaining self-tolerance, a transcriptome and an activation profile distinguishable from those of T regs produced in adults. This concept adds to the notion of a developmentally layered immune system. Other immune response-related genes as well as environmental factors probably play a role in development of the syndrome. Several studies of large cohorts of patients from different ethnic backgrounds have reported the appearance of chronic candidiasis at different sites in all patients. Moreover, the development of classic autoimmunity varies greatly between patients; however, the development of anticytokine-specific humoral immune responses appears to serve a more common characteristic of the disease. In one study, anti-interferon (IFN) antibodies in patients with APS-1 were detected in all Finnish and Norwegian patients with APS-1. The early finding of high-titer autoantibodies against IFNs has created a useful diagnostic test for APS-1. Neutralizing anti-IL-17 autoantibodies in patients with APS-1 have also been described. IL-17 is produced by a subset of T cells, T_{H17} cells, and plays an important part in the response to pathogens, particularly at mucosal surfaces. Additional analysis of IL-17 and IL-22 responses to the presence of Candida albicans antigens showed a marked deficiency in the ability of patients with APS-1 to produce IL-17 and IL-22. These findings raise an intriguing new aspect for the role of anticytokine autoantibodies,

which seem to provide a permissible environment for the development of chronic mucocutaneous candidiasis. Hypoparathyroidism and Addison disease present with similar high frequency (see Table 2–3). The occurrence of the diagnostic triad reportedly presents in 57% of patients. Female hypogonadism, presenting as total or partial failure of pubertal development or as premature ovarian failure in adults, has been reported in up to 60% of patients. Male hypogonadism is less frequent (14%). Type 1 diabetes is not as frequent as in APS-2, but if present, usually develops early (under 21 years of age). Autoimmune hypothyroidism (atrophic thyroiditis) is also less frequent than in APS-2; however, thyroid autoantibodies may be present in many euthyroid patients. Other manifestations are described in Table 2–3. Acute autoimmune hepatitis is reportedly less common than chronic hepatitis, which appears to be present in most individuals. Autoantibodies to aromatic L-amino acid decarboxylase (AADC) are associated with chronic active autoimmune hepatitis and vitiligo, which are found in APS-1. These antibodies, if present, can be helpful in making the diagnosis. Autoantibodies against tryptophan hydroxylase have been associated with gastrointestinal dysfunction in APS-1. Autoantibodies to the H^+-K^+-ATPase and to intrinsic factor are associated with pernicious anemia, and autoantibodies to tyrosinase are associated with vitiligo. Other autoantibodies associated with the single-gland disorders that make up this polyendocrine syndrome have been discussed above.

AUTOIMMUNE POLYENDOCRINE SYNDROME 2 (APS-2)

APS-2 is the most common of the polyendocrine failure syndromes. It affects women in a 3:1 ratio to men. APS-2 is diagnosed when at least two of the following are present: adrenal insufficiency, autoimmune thyroid disease (thyroiditis with hypothyroidism or Graves disease with hyperthyroidism), and type 1 diabetes. Historically, Schmidt (1926) first described the association of Addison disease and thyroiditis. Carpenter and coworkers in 1964 included type 1 diabetes in the syndrome. Other components of APS-2 include the following (see Table 2–3): primary hypogonadism, myasthenia gravis, celiac disease, pernicious anemia, alopecia, vitiligo, and serositis. The most frequent association appears to be with type 1 diabetes (over 50%) and autoimmune thyroid disease (70% in some series). Adrenal insufficiency may be concurrent, may be delayed in onset for up to two decades, or may never manifest. Some diabetic patients (2%-3%) develop celiac disease. Gluten-free diet is usually effective. If the celiac disease is untreated, hypocalcemia (not due to hypoparathyroidism), osteopenia, and occasionally gastrointestinal lymphoma may occur.

Although this syndrome and its components aggregate in families, there is no identifiable pattern of inheritance. Susceptibility is probably determined by multiple gene loci (HLA being the strongest) that interact with environmental factors. Many of the disorders of APS-2 are associated (some genetically linked) with the HLA haplotype identified in single organ disorders. HLA-A1, -B8, -DR3 and -DR4, DQA1*0501, and DQB1*0201 have all been described as associated with APS-2.

MANAGEMENT OF AUTOIMMUNE POLYENDOCRINE SYNDROMES

Hormonal replacement therapy remains the only form of treatment of the APS. The clinical management of these disorders mandates early diagnosis of associated components. Since the age at onset of associated disorders is clinically unpredictable, long-term follow-up is needed. Endocrine disorders are treated as they develop and are diagnosed. Hormonal treatments for the specific gland failures are described elsewhere in this book. However, specific combinations of endocrine organ failure require specific management. For example, thyroxine replacement can precipitate life-threatening adrenal failure in a patient with untreated Addison disease. Furthermore, hypoglycemia or decreasing insulin requirements in a patient with type 1 diabetes may be the earliest symptom/sign of adrenal insufficiency. Hypocalcemia, seen in APS-2, is more commonly due to celiac disease than hypoparathyroidism. Treatment of mucocutaneous candidiasis with ketoconazole in patients with APS-1 may induce adrenal insufficiency in a failing gland (ketoconazole is a global P450 cytochrome inhibitor). These antifungal drugs may also elevate liver enzymes, making the diagnosis of autoimmune hepatitis—requiring treatment with immunosuppressants—more difficult in these patients.

Screening of affected individuals as well as their relatives is the only way of preventing morbidity and mortality. Annual measurement of TSH is recommended as cost-effective in first-degree relatives of patients with type 1 diabetes. Autoantibody measurements may help in the preclinical assessment of several disorders. Complete blood counts, electrolytes, calcium and phosphorus levels, thyroid and liver function tests, blood smears (including RBC indices), and vitamin B_{12}/plasma methylmalonic acid measurements are all recommended in the follow-up of APS-1. For APS-2 patients with type 1 diabetes, thyroid disease and celiac disease coexist with sufficient frequency to justify not only TSH measurement but also screening for endomysial antibodies containing transglutaminase antibodies, which are prevalent in celiac disease.

IMMUNODEFICIENCY, POLYENDOCRINOPATHY, AND ENTEROPATHY, X-LINKED SYNDROME

Another autoimmune polyendocrine failure syndrome, immunodeficiency, polyendocrinopathy, and enteropathy, x-linked (IPEX) syndrome, is characterized by development of overwhelming systemic autoimmunity in the first year of life resulting in the observed triad of watery diarrhea, eczematous dermatitis, and endocrinopathy seen most commonly as type 1 diabetes mellitus. Most children have other autoimmune phenomena including Coombs positive anemia, autoimmune thrombocytopenia, autoimmune neutropenia, and tubular nephropathy. The majority of affected males die within the first year of life of either metabolic derangements or sepsis; a few survive into the second or third decade.

Diagnosis is based on clinical findings. FOXP3 is the only gene currently known to be associated with IPEX syndrome. Approximately 50% of males with IPEX syndrome have mutations identified in FOXP3. Genetic testing is clinically available. FOXP3 is expressed primarily in lymphoid tissues (thymus, spleen, and lymph nodes), particularly in CD4+ CD25+ regulatory T lymphocytes. In mice, it is required for the development and suppressive function of this important T reg population. In humans, it is not expressed at baseline in CD4+ CD25– or CD8+ T cells but is expressed upon T-cell activation. The FOXP3 protein is absent (due to nonsense, frameshift, or splicing mutations) in individuals with severe, early-onset IPEX syndrome. Some individuals with FOXP3 point mutations express a protein that appears to have decreased function, thereby leading to a milder form of the disease. Peripheral blood mononuclear cells from individuals with IPEX syndrome show an excess production of the T_{H2} cytokines IL-4, IL-5, IL-10, and IL-13 and decreased production of the T_{H1} cytokine IFN-γ.

Treatment options include: immunosuppressive agents (eg, cyclosporin A, tacrolimus) alone or in combination with steroids; sirolimus (rapamycin) for persons in whom tacrolimus therapy is toxic or ineffective; granulocyte colony-stimulating factor (G-CSF, filgrastim) for autoimmune neutropenia; nutritional support; and standard treatment of diabetes mellitus and autoimmune thyroid disease. If performed early, bone marrow transplantation (BMT) using nonmyeloablative conditioning regimens can resolve clinical symptoms. If the family-specific mutation is known, FOXP3 sequence analysis in at-risk males can be undertaken immediately after birth to permit early diagnosis and BMT before significant organ damage occurs; otherwise, monitoring at-risk males for symptoms is needed to enable early diagnosis and treatment.

IPEX syndrome is inherited in an x-linked manner. The risk to siblings of the proband depends on the carrier status of the mother. If the mother of the proband is a carrier, the chance of transmitting the disease-causing mutation in each pregnancy is 50%. Males who inherit the mutation will be affected; females who inherit the mutation are carriers and will not be affected. Affected males pass the disease-causing mutation to all of their daughters and none of their sons. Prenatal testing for pregnancies at risk is possible for families in which the disease-causing mutation has been identified.

POEMS SYNDROME (OSTEOSCLEROTIC MYELOMA)

The POEMS (polyneuropathy, organomegaly, endocrinopathy, M spike, skin abnormalities) syndrome, which is frequently seen as a concomitant of Castleman disease (giant cell lymph node hyperplasia), includes a variety of endocrinopathies of the adrenal, thyroid, pituitary, gonads, parathyroids, and pancreas.

The POEMS syndrome displays a number of endocrinopathies in the setting of lymphoproliferative disorders and presumed B-cell dysfunction. Aside from one report in which 2 of 11 patients with monoclonal gammopathy and some form of autoimmunity had POEMS, the endocrinological manifestations of

POEMS are not yet established as autoimmune in origin. When associated with Castleman disease, Kaposi-associated herpes virus (HHV8) may be implicated in the pathogenesis of the lymphoproliferation and the gammopathy.

Two-thirds of patients with POEMS reportedly had at least one endocrine abnormality at presentation. During the course of disease, endocrine abnormalities developed in another 10% of patients with POEMS. Hypogonadism seems to be the most common endocrine abnormality. Elevated levels of follicle-stimulating hormone (FSH) in the absence of primary hypogonadism have been reported. One-third of patients reportedly have erectile dysfunction with low serum testosterone levels. Fourteen percent of patients have hypothyroidism requiring therapy. An additional 12% had mild increases in TSH levels but normal thyroxine levels in one series. Sixteen percent of patients with POEMS have abnormalities of the adrenal-pituitary axis at presentation; with 5% of patients developing adrenal insufficiency later in the course of their disease. Three percent of patients have diabetes mellitus. Serum levels of parathyroid hormone were increased in three of four patients in whom it was measured in one series of 99 POEMS patients. Finally and although still hypothetical for POEMS, autoantibody-mediated mechanisms of disease (Graves disease) have been described in patients with other gammopathies.

REFERENCES

Akirav EM, Ruddle NH, Herold KC. The role of AIRE in human autoimmune disease. *Nat Rev Endocrinol*. 2011;7:25. [PMID: 21102544]

Brown EM. Anti-parathyroid and anti-calcium sensing receptor antibodies in autoimmune hypoparathyroidism. *Endocrinol Metab Clin North Am*. 2009;38:437. [PMID: 19328421]

Cooper GS, Bynum ML, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J Autoimmun*. 2009;33:197. [PMID: 19819109]

Holmes D. Thyroid gland. Unlocking the potential of thyroid autoantibodies. *Nat Rev Endocrinol*. 2014;10:314. [PMID: 24776731]

Imam S, Mirmira RG, Jaume JC. Eukaryotic translation initiation factor 5A inhibition alters physiopathology and immune responses in a "humanized" transgenic mouse model of type 1 diabetes. *Am J Physiol Endocrinol Metab*. 2014;306:E791. [PMID: 24496311]

Murphy K. *Janeway's Immunobiology*. 8th ed. New York: Garland Science; 2011.

Yang S, Fujikado N, Kolodin D, Benoist C, Mathis D. Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. *Science*. 2015;348:589. [PMID: 25791085]

This page intentionally left blank

Evidence-Based Endocrinology and Clinical Epidemiology

David C. Aron, MD, MS and Ajay Sood, MD

ARR	Absolute risk reduction		**NNH**	Number needed to harm
CI	Confidence interval		**PiD**	Positivity in disease
CV	Coefficient of variation		**QALY**	Quality-adjusted life year
EBM	Evidence-based medicine		**RCT**	Randomized controlled trial
ER	Event rate		**ROC**	Receiver-operating characteristic
NiH	Negativity in health		**RRR**	Relative risk reduction
NNT	Number needed to treat		**USPSTF**	U.S. Preventive Services Task Force

The individual practitioner faces a multiplicity of potential diagnoses, limitations in diagnostic capacity, subclinical disease identified by tests rather than by clinical manifestations, and rapid changes in scientific knowledge. The paradigm of clinical decision-making based on the assumption that all that is needed to guide clinical practice is personal experience (however unsystematic), understanding of pathophysiology, and thorough training plus common sense is insufficient to address these challenges. Moreover, the integration of relevant research findings into clinical practice has been haphazard; the lag time between development of scientific knowledge and introduction into practice can be many years, and there is marked variation in practice. A systematic approach based on principles of clinical epidemiology can help address some of these issues. This quantitative approach has formed the primary, albeit not the only, basis of the evidence-based medicine (EBM) movement. This movement posits that understanding certain rules of evidence is necessary to interpret the literature correctly and that physicians who practice based on the above understanding will provide superior care. This chapter summarizes some of the principles of clinical epidemiology and evidence-based endocrinology and some of their limitations.

CLINICAL EPIDEMIOLOGY

Clinical epidemiology consists of the application of epidemiologic principles and methods to problems encountered in clinical medicine. Clinical epidemiology emphasizes a quantitative approach and is therefore concerned with counts of clinical events. Its applications are paramount (1) in diagnostic testing and how the results modify the probability of a particular disorder being present and (2) in treatment decisions in which the potential benefits and harms must be addressed. The techniques of clinical epidemiology have become increasingly important as practitioners confront the complexity of contemporary medical practice.

DIAGNOSTIC TESTING: TEST CHARACTERISTICS

The appropriate choice and interpretation of diagnostic tests, whether biochemical assays, radiologic procedures, or clinical findings, have always been essential to the clinical practice of endocrinology. These tests, when introduced in the medical literature, are accompanied by varying degrees of validation. The clinician's

assessment of the utility of tests, now even more important with the emphasis on cost-effectiveness, can be improved by knowledge of test EBM principles. We review some of these concepts as they apply to diagnosis and management of endocrine disorders, including the topics of test characteristics such as sensitivity and specificity, receiver-operating characteristic (ROC) curves, likelihood ratios, predictive values, and diagnostic accuracy.

The evaluation of endocrine function begins with a clinical question. The more vague the question, the more difficult it is to obtain a clear answer. Part of this step involves a clinical judgment about the likelihood of the disease prior to obtaining a test and its results. This pretest probability is combined with the performance characteristics of the test and its use (sensitivity and specificity, ROC curves, likelihood ratios, predictive values, and diagnostic accuracy) in order for proper interpretation. Variation is inherent in biological systems. Therefore, diagnostic tests must take into account not only variability of the tests themselves and how they are performed but also variability in the populations in whom the tests were developed, both with the disease and without the disease. Key aspects in the analysis of a test include reproducibility (precision) and accuracy. **Reproducibility** describes how close the test comes to producing the same results every time and depends on such factors as intraobserver and interobserver variability (as in the assessment of a physical finding or x-ray) and, in the case of biochemical tests, characteristics such as intra-assay and interassay coefficients of variation (CVs). Although studies utilizing radioimmunoassays and other assays routinely report intra-assay and interassay CVs, few papers publish multiple results performed on the same patient (intra-individual variation). There have also been relatively few studies on the reliability of measurements (ie, the degree of intra-individual variation). One study from 1994 found that the minimum number of replicate measurements necessary to achieve satisfactory reliability of the mean of basal levels was 3 for plasma cortisol and 18 for salivary cortisol. Responses to dynamic tests required fewer replicates to achieve the same reliability (one to two samples).

Reproducibility depends on the conditions under which the test is performed. Clinicians must be aware of the distinction between efficacy and effectiveness when translating published results into practice. As applied to diagnostic testing, efficacy refers to the degree to which the test has been shown scientifically to accomplish the desired outcome. In contrast, effectiveness refers to the degree to which the test achieves this outcome in actual clinical practice. Most large studies have been performed in research venues and thus are efficacy studies, whereas the effectiveness of most tests in practice has not been extensively evaluated. In comparing one's own results with a published report or laboratory normal range, it is important to take into account those conditions (eg, test performed in a hospital vs a physician's office).

Accuracy describes how close the test comes to producing results that are a true measure of the phenomenon of interest; systematic bias of a highly reproducible test may produce the same incorrect result every time. Like reproducibility, accuracy depends on the conditions under which the test is performed; accuracy in the clinical practice setting may differ from that in the experimental setting where many extraneous influences are controlled.

When interpreting a test, the result is usually compared to a normal range. Frequently, the normal range is developed by using a reference population assumed (or preferably shown) to be disease free. For example, when designing a test to be used in the diagnosis of Cushing syndrome, the reference group should be made up of individuals who have clinical features suggestive of Cushing syndrome but who in fact do not have the disorder. However, reference groups may be made up of individuals who are readily accessible and not more appropriate comparisons. It is also important to note that in establishing a normal range based on a Gaussian or normal distribution, encompassing the mean ± two standard deviations, 5% of disease-free individuals have a result outside the limits of normal. (It is important to note that the definition of normal based on a Gaussian distribution of values is only one of a number of definitions of normal. Some others include the one most representative or most common in a class, the one most suited to survival, and the one that carries no penalty [ie, no risk].) Figure 3–1 illustrates a normal range and a range in a population with disease. A result outside normal limits is not equivalent to disease. Moreover, values within the normal range do not necessarily exclude disease. Values in the population of individuals with disease are determined separately, and the overlap with the normal range is assessed.

Sensitivity and Specificity

Ideally, a diagnostic test has no overlap between the results of individuals with the disease and those without the disease. The reality, however, is different. Test characteristics that describe this overlap are sensitivity and specificity, and they are typically illustrated in a 2 × 2 table. As shown in Figure 3–2, sensitivity and specificity are collectively known as **operating characteristics**. Sensitivity refers to the ability to identify correctly individuals with a disease. The sensitivity of a sign or symptom (or diagnostic test) is the proportion of patients with disease who has a positive test result, sign, or symptom. In contrast, specificity refers to the ability to identify correctly individuals without a disease. The specificity of a test is the proportion of healthy patients who have a negative test result or who lack that sign or symptom.

Thus, the **sensitivity** of a test equals the number of individuals with disease who have a positive test (true positive [TP]) divided by the number of individuals with disease (true positives plus false negatives [FN]), whereas **specificity** equals the number of individuals without the disease who have a negative test (true negatives) divided by the number of individuals without disease (true negatives plus false positives [FP]). Sensitivity is sometimes termed **PiD** or positivity in disease, and specificity is sometimes termed **NiH** or negativity in health. In theory, sensitivity and specificity are characteristics of the test itself and not of the patients on whom the test is applied. However, this may not be correct in practice. The sensitivity of a test may be affected by the stage or severity of the disease. The specificity of a test may depend on the characteristics of the reference population. The nature of the groups used to establish the cut-points that differentiate normal from abnormal must be appropriate and should be specified in any report of a diagnostic test. The value chosen for a cutoff point
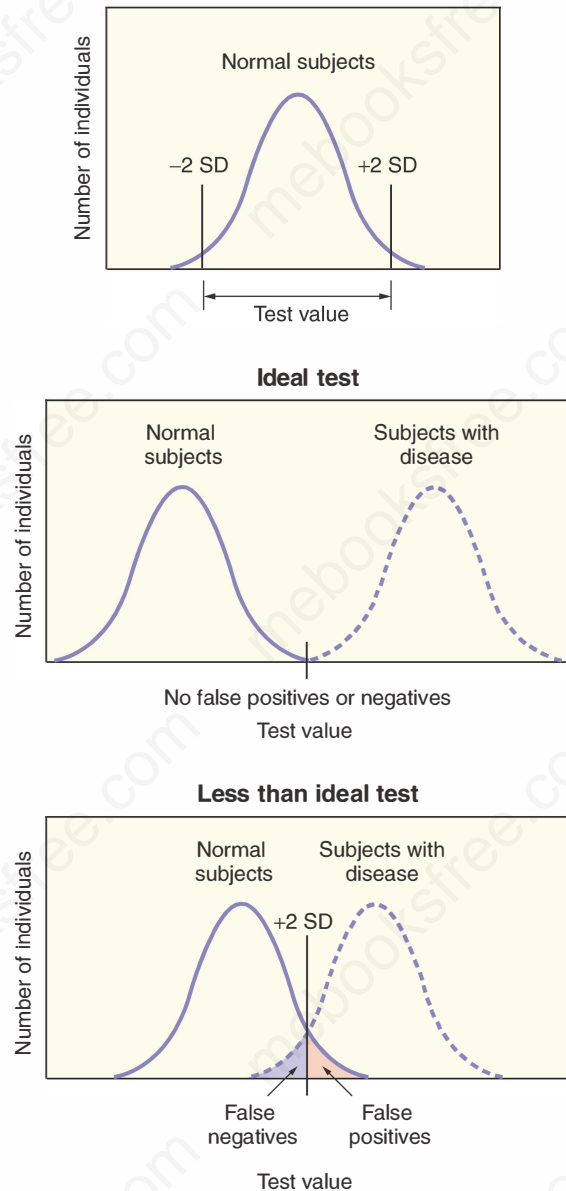
FIGURE 3–1 Defining a normal range and diagnostic testing. The top panel shows the Gaussian (normal) distribution of test values in a population of individuals. The middle panel illustrates two such curves that describe the findings in an ideal test. In this case, there is no overlap in the results between normal subjects and subjects with disease (ie, there are no false-positive results and no false-negative results). The bottom panel illustrates the results for a less-than-ideal test. Normal subjects with test values above the given threshold have abnormal results (ie, false-positive results), whereas some subjects with disease have test values below that threshold (ie, normal or false-negative results).

also affects the sensitivity and specificity. To assist the clinician in assessing a report on a diagnostic test, a series of questions has been proposed (Table 3–1).

The overnight dexamethasone suppression test is commonly used as a screening test in the diagnosis of Cushing syndrome, and its use illustrates some of the issues in diagnostic testing. As shown

in Figure 3–3, combining the results of a number of studies indicates a sensitivity of 98.1% and specificity of 98.9%. However, the individual studies varied, with sensitivities ranging from 83% to 100% and specificities from 77% to 100%. Moreover, the studies used different cortisol assays, doses of dexamethasone, and criteria for a positive test. Caution must be exercised in drawing conclusions from the combining of such data. In order to apply the sensitivity and specificity of a test derived from one study sample to a different population, the test cannot deviate from methods used (eg, dose of dexamethasone, type of cortisol assay, and timing of dexamethasone administration) when the optimal cutoff was determined, and the sample studied must be similar to the new population to be tested. To meet this latter prerequisite, the sample studied must account for variability of diseased individuals. This requires that subjects with disease be defined using the best available gold standard (independent of the test in question) and include a broad enough cross-section of those with disease (eg, mild vs severe disease, different etiologies of disease, as well as age, sex, and race) to establish a reliable range of measurements. The characteristics of the reference sample of subjects without the disease are equally important. Although the 1-mg overnight dexamethasone suppression test is still believed to have excellent, albeit less than 100% sensitivity, it has serious problems with specificity, and false-positive results have been described with a variety of drugs as well as medical, surgical, and psychiatric conditions.

One additional method of reporting the performance of a test is diagnostic accuracy. This can also be derived from the 2 × 2 table. Diagnostic accuracy is defined as the ability of a test to identify correctly those with the disease and those without the disease:

$$\text{Diagnostic accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

From Crapo's composite data on the 1-mg overnight dexamethasone suppression test, the diagnostic accuracy is calculated as 98.7%. In addition to the characteristics of the study subjects, the number of subjects included in the sample is also critical to assess the accuracy of the test. Each operating characteristic determined in a study should be accompanied by a **confidence interval (CI)**—a range of values calculated from sample size and standard error which expresses a degree of confidence (usually 90%, 95%, or 99%) that the unknown true sensitivity or specificity lies in that interval. CIs are a measure of the precision of an estimate. The range of a CI depends on two factors: (1) the number of observations; and (2) the spread in the data (commonly measured as a standard deviation). The fewer the number of observations, the larger the range of the CI, and the greater the standard deviation of the data, the larger the range of the CI.

In addition to the limitations on the operating characteristics based on the samples from which the data are derived, sensitivity and specificity are not independent of each other. They vary with the cutoff level chosen to represent positive and negative test results. In general, as sensitivity increases, specificity decreases and as specificity increases, sensitivity decreases. This phenomenon is depicted graphically in an ROC curve.

		Disease	
		+ Present	**− Present**
Test result	+	a True positive (TP)	b False positive (FP)
	−	c False negative (FN)	d True negative (TN)

Sensitivity = a/(a + c) = TP/(TP + FN)

Specificity = d/(b + d) = TN/(TN + FP)

Positive predictive value = a/(a + b) = (TP/TP + FP)

Negative predictive value = d/(c + d) = TN/(TN + FN)

Diagnostic accuracy = (a + d)/(a + b + c + d) = (TP+TN)/(TP + FP + TN + FN)

Pretest (prior) probability = (a + c)/(a + b + c + d) = (TP + FN)/(TP + FP + TN + FN)

Likelihood ratio for a positive test result (LR+):

    = [a/(a + c)]/[b/b(b + d)] = TP rate/FP rate = [TP/(TP + FN)/(FP/TN + FP)]

    = sensitivity/(1 − specificity)

Likelihood ratio for a negative test result (LR−):

    = [c/(a + c)]/[d/(b + d)] = FN rate/TN rate = [FN/(TP + FN)]/[TN + FP]

    = (1 − sensitivity)/specificity

FIGURE 3–2 This 2 × 2 table graphically depicts how tests can be performed in individuals with and without a disease.

ROC Curves

An ROC curve graphically illustrates the trade-off between the false-negative and false-positive rates for different cutoff points of a diagnostic test. In an ROC curve, the true-positive rate (sensitivity) is plotted on the vertical axis and the false-positive rate (1–specificity) is plotted on the horizontal axis for different cutoff points for the test. The dotted diagonal line in Figure 3–4 corresponds to a test that is positive or negative just by chance (ie, the true-positive rate equals the false-positive rate). Such a test provides no useful information. Ideally, a test would provide results that could be plotted on one point in the top left corner—100%

true-positive rate and 100% true-negative rate. The closer an ROC curve is to the upper left-hand corner of the graph, the more accurate it is, because the true-positive rate is 1 and the false-positive rate is 0. As the criterion for a positive test becomes more stringent, the point on the curve corresponding to sensitivity and specificity (point A) moves down and to the left (lower sensitivity, higher specificity); if less evidence is required for a positive test, the point on the curve corresponding to sensitivity and specificity (point B) moves up and to the right (higher sensitivity, lower specificity). Analysis of the area between the actual results and the straight line indicates how good the test is. The greater the area under the curve, the better the test.

Depending on the purpose of the test, the curves may be used to decide an optimal cutoff level for a single test. For example, with a screening test, high sensitivity is typically desired, and the trade-off is lower specificity. The cutoff point may also be chosen depending on health costs (morbidity and mortality associated with an error in diagnosis), financial costs, or need for maximal information (the operating position giving the greatest increase in posttest probability).

ROC curves may also be used to compare two or more tests by comparing the areas under the curves, which represent the inherent accuracy of each test. An example of the comparison of the performance of different tests for the diagnosis of pheochromocytoma is given in Figure 3–5. It is important to remember, however, that ROC curves are only as good as the operating characteristics from which they are generated.

Finally, determining cost-effective diagnostic strategies requires careful evaluation not only of a test in isolation, but also in the context of the other information available and the likelihood of disease. This is the essence of Bayesian models of decision-making.

TABLE 3–1 Guide for evaluating studies of test performance.

Question #1	Is the study population described well enough?
Question #2	Does the spectrum of illness in the study population differ from that of my patients (eg, spectrum bias)?
Question #3	Was a positive result on the index test a criterion for referral to have the gold standard test?
Question #4	Was there blinding of those who interpreted the index test and those who interpreted the gold standard test (eg, test review bias)?
Question #5	Was the gold standard test an adequate measure of the true state of the patient?

Data from Jaeschke R, Guyatt G, Sackett DL. Users' guide to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA.* 1994;271:389-391; and Jaeschke R, Guyatt G, Sackett DL. Users' guide to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA.* 1994:271:703.

		Cushing syndrome	
		Present	Absent
1-mg overnight dexamethasone suppression	No suppression	151	5
	Suppression	3	461

Sensitivity = 151/(151 + 3) = 98.1%

Specificity = 461/(5 + 461) = 98.9%

A

		Cushing syndrome	
		Present	Absent
1-mg overnight dexamethasone suppression	No suppression	151	101
	Suppression	3	858

Specificity = 858/(101 + 858) = 89.5%

B

		Cushing syndrome	
		Present	Absent
1-mg overnight dexamethasone suppression	No suppression	151	96
	Suppression	3	397

Specificity = 397/(96 + 397) = 80.5%

C

FIGURE 3–3 Diagnosis of Cushing syndrome with the 1-mg overnight dexamethasone suppression test: test characteristics with normal controls (**Panel A**); all controls (**Panel B**); and "obese" and "other" controls (**Panel C**). These data show how the specificity of the test is affected by the types of control subjects. (Data from Crapo L. Cushing's syndrome: a review of diagnostic tests. *Metabolism*. 1979 Sep;28(9):955-977.)

In this model, the physician updates his or her belief in a hypothesis with each new item of information, with different weights given to the new information depending on its operating characteristics. Consideration must be given to the question of the value added by a test or procedure. This can be assessed with ROC curves and statistical models.

Predictive Values, Likelihood Ratios, and Diagnostic Accuracy

Sensitivity and specificity are important test characteristics, yet the clinician wants to know how to interpret a test result. Predictive values help in this regard. As shown in Figure 3–6, the **positive predictive value** is the proportion of patients with a positive test who actually have the disease. Similarly, the **negative predictive value** is the proportion of those with a negative test who do not have the disease. Because each of these values are calculated using results from both individuals with and without the disease in question, the prevalence of the disease has a great impact on the values. For any given sensitivity and specificity, the lower the prevalence

of disease (or the lower the pretest probability), the more false-positive results there are (see Figure 3–6).

The **likelihood ratio**, which is derived from sensitivity and specificity, is an expression of the odds that a sign, symptom, or test result is expected in a patient with a given disease as opposed to one without the disease. Two forms of the likelihood ratio exist, the likelihood ratio for a positive finding and the likelihood ratio for a negative finding. Calculations are shown in Figure 3–2. Likelihood ratios offer some advantages over sensitivity and specificity. They are the most useful in calculating posttest probabilities given prevalence (a probability) and likelihood ratios. A convenient nomogram for this has been published (Figure 3–7).

AN APPROACH TO DIAGNOSIS IN PRACTICE

In approaching a case, the clinician makes a series of inferences about the nature of a patient's condition and proceeds toward a working diagnosis along with some alternatives—generation of a
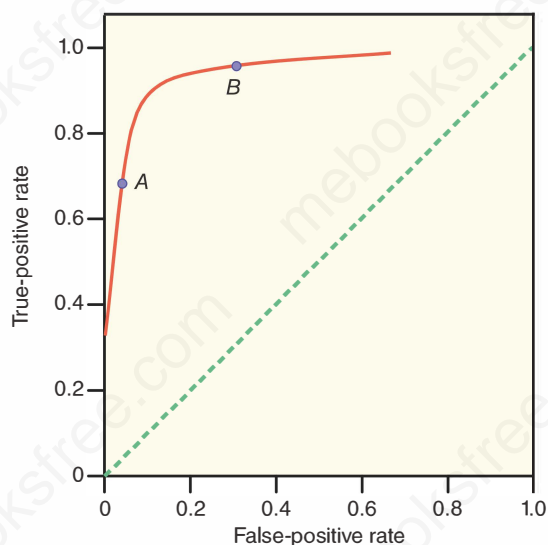
FIGURE 3–4 Receiver-operating characteristic (ROC) curve. In an ROC curve, the true-positive rate (sensitivity) is plotted on the vertical axis, and the false-positive rate (1 – specificity) is plotted on the horizontal axis for different cutoff points for the test. The dotted diagonal line corresponds to a test that is positive or negative just by chance (ie, the true-positive rate equals the false-positive rate). The closer an ROC curve is to the upper left-hand corner of the graph, the more accurate it is, because the true-positive rate is 1 and the false-positive rate is 0. As the criterion for a positive test becomes more stringent, the point on the curve corresponding to sensitivity and specificity (**point A**) moves down and to the left (lower sensitivity, higher specificity); if less evidence is required for a positive test, the point on the curve corresponding to sensitivity and specificity (**point B**) moves up and to the right (higher sensitivity, lower specificity). Analysis of the area between the actual results and the straight line indicates how good the test is. The greater the area under the curve, the better the test.

differential diagnosis. Although one could create a listing of all possible causes of the patient's complaint (ie, a possibilistic differential diagnosis), experienced clinicians generate a differential diagnosis that is a combination of probabilistic (ie, considering first those disorders that are more likely), prognostic (ie, considering first those disorders that are more serious if missed), and pragmatic (ie, considering first those disorders that are most responsive to treatment). The clinician then refines the diagnostic hypotheses, sometimes using clues from the history and physical examination and often with the use of laboratory or radiologic tests. In so doing, the clinician tries to reduce the inherent uncertainty so that the most appropriate course of treatment can be prescribed.

The first step in this process is to understand the concept of probability. A probability is an expression of likelihood and thus represents an opinion of the relative frequency with which an event is likely to occur. In the case of diagnosis, probability is a numerical expression of the clinician's uncertainty about a diagnosis; expressing a clinical opinion in subjective terms such as *likely* and *possible* is fraught with imprecision and misunderstanding. Certainty that a disease is present is assigned a probability of one,
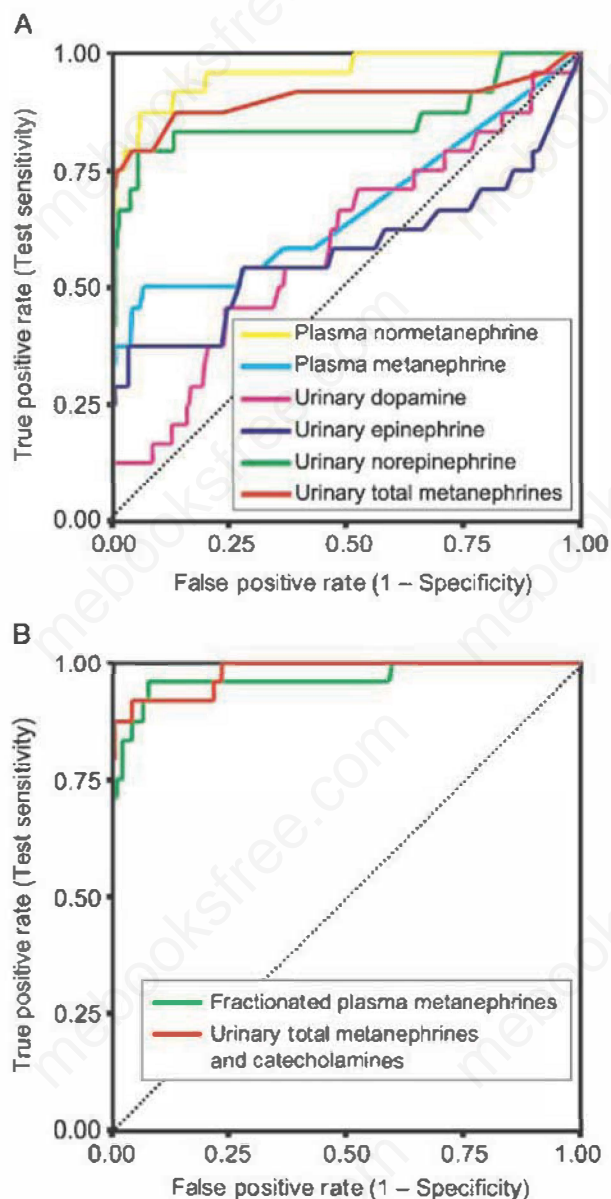
FIGURE 3–5 Receiver-operating characteristic (ROC) curves for diagnostic tests for pheochromocytoma. The area under the curve from urinary total metanephrines and catecholamines and fractionated plasma metanephrines in Figure B exceeds that of other tests in Figure A indicating that they are more accurate overall.

Disease prevalence or pretest probability	0.1%	1%	10%	50%	90%
Positive predictive value	0.89%	8.33%	50.0%	90.0%	98.78%
Negative predictive value	99.99%	99.89%	98.78%	90%	50.0%

FIGURE 3–6 Positive and negative predictive values as a function of disease prevalence, assuming test sensitivity and specificity of 90% for each.
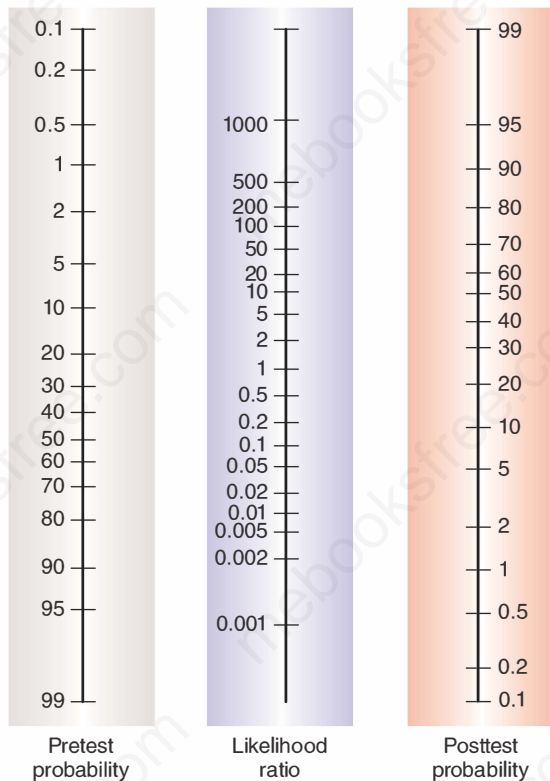
FIGURE 3–7 Nomogram for likelihood ratios. (Modified with permission from Fagan T J. Letter: Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul 31;293(5):257.)

certainty that a disease is not present is assigned a probability of zero, and a clinician's opinion of the disease being present or absent usually falls somewhere in between. Of course, probabilities are derived from different data sources that vary in their reliability and application to a given patient, such as the clinician's experience (remembered cases), research studies, and population-based epidemiologic studies. Therefore, some degree of uncertainty is inherent in a given probability, and the confidence with which one can rely on a given probability depends to a large extent on the data underlying it.

The diagnostic approach to minimizing uncertainty requires four steps. First, the clinician starts with an estimate of probability based on initial clinical impressions. This starting point is dubbed the *pretest* or a priori probability and is a number between zero and one that expresses the likelihood of disease. For example, a clinician who sees a large population of patients with diabetes mellitus may think that a 55-year-old patient with polyuria, polydipsia, weight loss, and fatigue has a probability of 0.70 of having diabetes mellitus (ie, if there were 100 such individuals, 70 would have diabetes).

Second, the clinician determines the *threshold probability for treatment*. The treatment threshold probability is defined as the probability of disease at which one is indifferent between giving treatment and withholding treatment. Establishing a treatment threshold probability takes into account the costs (not just in the monetary sense) and benefits of treating or not treating.

Because probability is predicated on the lack of certainty about the presence or absence of disease, it is inevitable that some patients who are not diseased receive treatment and others who are diseased do not receive treatment. Third, if the pretest probability is greater than the threshold probability, the clinician chooses to treat; if it is less than the threshold probability, the clinician opts not to treat. If the clinician is not comfortable enough about the presence or absence of disease, he or she may choose to order further tests with the objective of getting closer to certainty. The fourth step involves taking information gained from the test and using it to update the pretest probability. The updated, or *posttest*, probability can also serve as a new *pretest* probability for the next step in hypothesis testing (Figure 3–8).
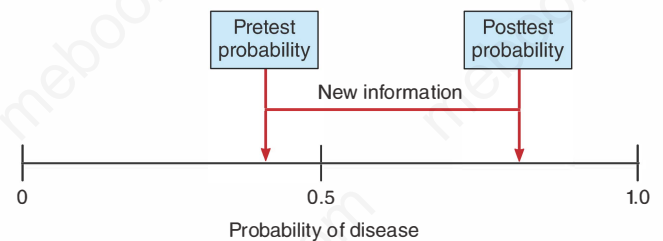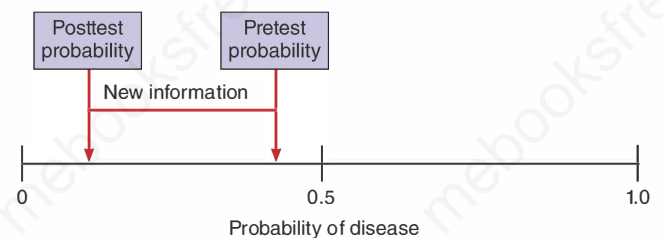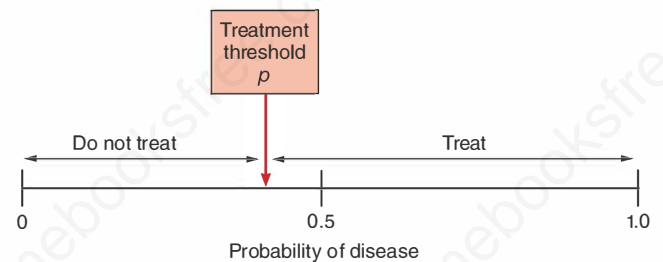
FIGURE 3–8 Adjusting probabilities with new information and treatment thresholds. The top panel shows a pretest probability of disease of approximately 0.4. With the new information provided by a test, the probability rose to approximately 0.7. The middle panel shows the same pretest probability, but a negative test result reduced the probability of disease to approximately 0.15. The treatment threshold probability is the probability above which one would treat. The actual threshold value depends on the morbidity and mortality of disease and the adverse effects (morbidity and mortality) of treatment.

Treatment threshold probability depends on the effectiveness of treatment, including its adverse effects. In general, treatment threshold probability is generally low when treatment has a high benefit for diseased patients and/or a low risk of harm for non-diseased patients. Likewise, the treatment threshold probability is generally high when there is a high risk to nondiseased patients and/or a low benefit to diseased patients. For example, an endocrinologist evaluating an incidentally found pituitary microadenoma would need to be as certain as possible that the patient suffered from hormone hypersecretion before recommending surgical resection because of the high risk of harm to nondiseased patients relative to the benefit to diseased patients. The same principles apply to diagnostic testing. There is a *test-threshold* defined as the probability of disease at which one is indifferent between performing a test and not performing a test. Depending on the circumstances, one would treat without performing the test or not perform the test and ignore the issue entirely.

Tests can be combined in the hope that diagnostic accuracy is enhanced. Two tests can be performed in parallel (simultaneously) or in series (sequentially). When two tests are performed in parallel, a positive result in either test establishes the diagnosis; when two tests are performed in series, positive results in both tests are required to make the diagnosis. Thus, parallel testing increases sensitivity at the cost of specificity, whereas series testing increases specificity at the cost of sensitivity. For this approach to be better than one test alone, the second test has to provide information not provided by the first test. A common practice is to repeat the same test. This approach may be quite logical when the disorder is intermittently active (eg, intermittent Cushing syndrome) or when there is concern about the way the first test was performed. When tests are performed in series, the first test is usually the one with the higher sensitivity so that as many patients with the disease are discovered. False-positive results are then identified by a second test with higher specificity. Another approach to sequential testing uses as the first test the one with highest specificity so that fewer patients without the disease go on to further testing. This approach makes sense when the second test is associated with more morbidity. Adjusting cutpoints for the first test can make it 100% specific (at the cost of some sensitivity). This approach has been used frequently in tests related to Cushing syndrome. One caveat applies to the method of choosing the cutoff point. In addition to considering whether the patients without the disease are an appropriate control group, attention should be paid to sample size. Caution must be exercised when applying test characteristics derived from small samples.

CLINICAL EPIDEMIOLOGIC PRINCIPLES APPLIED TO TREATMENT DECISIONS

Critical to decision-making is the appropriate interpretation of the results of clinical trials and the impact of treatments on clinical events. The use of a 2 × 2 table can facilitate this process (Figure 3–9).

The **event rate (ER)** is the rate of occurrence of the outcome of interest. The **relative risk reduction (RRR)** is the proportional reduction in rates of bad outcomes between experimental and control participants in a trial, and it can be calculated as:

$$RRR = 1 - \text{relative risk} = 1 - (RRa/RRb)$$

RRR is usually reported accompanied by a 95% CI. **Absolute risk reduction (ARR)** is the absolute arithmetic difference in rates of poor outcomes between experimental and control participants in a trial and is usually accompanied by a 95% CI. Another way to think about this is to ask how many patients would have to be treated in order to prevent one negative outcome or achieve one additional favorable outcome. This is termed **number needed to treat (NNT)**, and it is calculated as 1/ARR and accompanied by a 95% CI. An example is provided (see Figure 3–9) from the results of a clinical trial. In the United Kingdom Prospective Diabetes Study (UKPDS-38), hypertensive patients with type 2 diabetes were randomized to tight versus less tight blood pressure control. The patients with tight blood pressure control had fewer diabetes-related complications and fewer deaths due to diabetes compared to the patients with less tight blood pressure control. Over a period of 8 years of treatment, 15.9% of patients in the less tight control group had died of diabetes complications compared with 10.9% in the tight control group. If intervention results in an adverse event, then an equivalent term to quantify that effect would be **number needed to harm (NNH)**, which would indicate that how many patients need to be treated to cause harm to one patient.

To apply these statistics in clinical decision-making to patients, it is necessary to calculate ARR, RRR, and NNT. The RRR of 32% indicates that approximately one-third of the expected deaths were prevented by tight control of blood pressure. However, although the RRR may be similar across different risk groups, the ARRs are not. ARRs become smaller when event rates are low, whereas RRR remains constant. Thus, RRR can be misleading. The lower the ER in the control group, the larger the difference between RRR and ARR. In other words, the benefits of treatment as reflected in ARR are usually greater in patients at higher risk of negative outcomes than in patients at lower risk of negative outcomes.

The RRR fails to discriminate absolute treatment effects that are clinically significant from those that are trivial. The NNT is 20, so that one would need to treat 20 patients for 8 years with tight blood pressure control to prevent one death from diabetes. NNT can be readily used to compare different treatments. In this case, RRR was 32% and ARR was 5.1%. If a different treatment with the same RRR of 32% has an ARR of 0.51%, the NNT would be 196, indicating that a much greater effort would be needed to achieve one additional favorable outcome. This might or might not be worth the effort, depending on a variety of factors. One method by which treatment thresholds can be determined is with the method of decision analysis.

		Outcome	
		Negative outcome	Positive outcome
Treatment group	Experimental	a	b
	Control	c	d

Relative risks

RRa—risk rate of the negative outcome occurring in the experimental group = a/(a + b)

RRb—risk rate of the negative outcome occurring in the control group = c/(c + d)

RR—the relative risk of the negative outcome occurring in experimental treatment compared to control treatment = RRa/RRb = {a/(a + b)}/{c/(c + d)}

Relative risk reduction (RRR) = (1 − RR) X 100%

Absolute risk reduction (ARR) = RRb − RRa

Number needed to treat (NNT) = 1/ARR

EXAMPLE:

Figure 2 X 2 Table for UKPDS Trial of Tight Control of Blood Pressure in Patients with Diabetes

		Outcome—Deaths from diabetes	
		Negative outcome —Death	Positive outcome —Alive
Treatment group	Experimental (n = 758)	82	676
	Control (n = 390)	62	328

RRa—10.8%
RRb—15.9%
RR—0.68
RRR—32%
ARR—5.1%
NNT—20

FIGURE 3–9 A 2 × 2 table for treatment groups and outcomes.

DECISION ANALYSIS

Decision analysis is a mathematical tool designed to facilitate complex clinical decisions in which many variables must be considered simultaneously. This analytical procedure selects among available diagnostic or therapeutic options based on the probability and predetermined value (utility) of all possible outcomes of those options. Decision analysis provides a systematic framework for organizing all data relevant to the decision so that relevant uncertainties are less likely to be overlooked. Performing a decision analysis requires clear definition of the relationship between possible courses of action and their associated outcomes and assignment of numerical values to various courses of action. In so doing, decision analysis simplifies comparisons among different strategies.

There is substantial variation in duration and severity of disease between individuals. Choosing a treatment option in the setting of unpredictable effects is a difficult problem, and expected value decision-making is a useful tool. When individual outcomes are uncertain, expected value is the result that is expected *on average*. In preparing to perform a decision analysis, one first must define the problem with a clear statement of the strategies to be examined. Two or more strategies may be included. Clinical decisions have effects over different time frames. An important step in defining a decision problem is to decide the time horizon for outcomes to be evaluated. For example, if strategies for diabetes treatment are to be compared, a period of weeks to months may be appropriate for evaluating the risk of hypoglycemic episode. However, this relatively short time horizon would not be appropriate when the study outcome is diabetic retinopathy, where a time horizon of years would be a better choice. After choosing the problem and a single time horizon, carrying out a formal analysis typically involves six general steps:

1. Construction of a decision tree that maps out all the possibilities
2. Determination and assignment of probabilities
3. Assignment of utilities to each potential outcome

4. Determination of the expected utility
5. Choosing the course of action with the highest expected utility
6. Evaluation of the sensitivity of the chosen course of action to changes in probabilities and utilities

In a *decision tree*, the term decision alternative refers to one of the potential strategies to be analyzed. Each alternative should be listed. Figure 3–10 shows a sample decision tree that outlines strategies of medical versus surgical treatment for disease. The decision itself is represented by a box called a decision node. All of the possible outcomes for each decision alternative are listed. An event that has outcomes under the control of chance is denoted by a chance node. The symbol for a chance node is a circle. The series of events leading to the clinical outcomes is represented by a series of chance nodes and decision nodes. The decision tree is usually written from left to right, with the initial decision node on the far left and the final outcomes on the far right. A final outcome is represented by a terminal node, the symbol for which is a triangle. There may be any number of outcomes at a chance node. The listed outcomes should include all possible outcomes, and they must not overlap. In addition to this assumption of

mutual exclusivity, structuring a tree in this fashion assumes that the probability of occurrence of one event does not influence the probability of occurrence of another event(s). The decision tree structure should be as similar as possible for all strategies, because differences may lead to a structural bias in the analysis.

The decision tree described earlier is relatively simple, and this approach cannot readily represent complex, dynamic clinical situations with recurring events. More elaborate models, such as Markov models, can be used to represent the passage through multiple health states. Patients may transition from one health state to another with some probability within a specified time period or model cycle. Like each terminal node in a static decision tree, each health state in the Markov model is associated with a specific clinical measure, utility, or cost. A key assumption of the Markov model is that the future is determined only by the individual's present health state; events prior to that health state or how long it took to arrive there do not affect the individual's future, a simplifying assumption that may not hold true for some health problems.

One strength of the decision analysis process is that it may be used for a variety of outcome measures. The outcome measure of interest determines the information needed for analysis. For example, one may use clinical measures such as survival following total thyroidectomy for differentiated thyroid cancer, preservation of vision after laser photocoagulation, or meeting a target level for glycosylated hemoglobin. Economic measures provide measures of cost and resource use. When cost measures are included, it is important to consider whose perspective is represented; analyses may reflect the viewpoint of society as a whole, payers, or patients.

A detailed analysis of costs may be used for *cost-effectiveness analysis or cost–utility analysis*. Utility measures reflect outcome preferences. For an individual, utilities represent quality of life preferences. They are quantitative values used to summarize multiple dimensions, which may be conflicting. For example, decision-makers frequently must choose among strategies that have differing effects on length of life and quality of life and involving trade-offs between the two. Utilities are traditionally scored on a scale from 0 to 1. The ideal situation, often perfect health, is scored as a 1. The worst situation, death, is scored as a 0, and this assumes that there are no utilities worse than death. Intermediate states are assigned values between 0 and 1. For example, living with diabetes mellitus may receive a score of 0.80—less than perfect, but preferable to death. Among the assumptions underlying this approach are: stability of estimates of utilities, that is, the estimates obtained prior to experiencing an event will not change after the event has been experienced; and the it-does-not-matter-how-you-get-there assumption (ie, death or other outcomes have the same utility regardless of the route a patient takes in getting there).

There are different ways to obtain utility values for different health states. One relatively simple method is the visual analog scale, in which a subject is asked to rate a given health state on a scale from 0 to 100. There also are more specific ways of determining utility. The standard gamble approach was developed by von Neumann and Morgenstern as a method for assessing utility.

FIGURE 3–10 A decision tree with probabilities and utilities. This sample decision tree outlines two strategies for a hypothetical patient with an adrenal mass—observation versus surgery. The decision node (open square) represents the decision itself. All of the possible outcomes for each decision alternative are listed emanating from an open circle or chance node. The events are governed by chance. The series of events leading to the clinical outcomes is represented by a series of chance nodes and decision nodes. Terminal nodes (triangles) denote the outcomes. Probabilities and utilities are listed. These are derived from the literature or determined via utility elicitation methods. In this example, if observation is chosen, one lives with the anxiety of uncertainty about the nature of the mass, which has a utility of 0.99 and would occur 95% of the time, while the health state of long-term consequences which occurs 5% of the time has a utility of 0.6. Similarly, surgery is curative 99% of the time, but has a 1% mortality rate. Cure, the most preferred outcome has a utility of 1 while death has a utility of 0. The predicted outcome of observation is $(0.99 \times 0.95) + (0.6 \times 0.05)$ or 0.9705 compared to $(1 \times 0.99) + (0 \times 0.01)$ or 0.99 for surgery. This indicates that on the *average*, surgery is a better choice in this particular case. However, if the utilities were different or the probabilities of the outcomes were different, the choice might also be different.

One advantage of this method is that it incorporates the participant's attitudes about risk taking, because the process involves consideration of a hypothetical gamble. For example, consider a 70-year-old individual with heart failure, diabetes, and chronic renal failure who is found to have a 4-cm adrenal mass on a CT scan performed for unrelated reasons. There is no indication of hormonal production by the mass, but knowing about this mass causes anxiety because it could be malignant. This concern could be addressed by surgical removal of the mass, but there is a risk of surgical complications. The individual is asked whether s/he would rather live with the mass (and the anxiety) or gamble on surgery that would eliminate the mass (curative), but has a 10% mortality rate. If the patient prefers to live with the mass, then ask would s/he prefer to live with the mass or gamble on curative surgery that has a 0.1% mortality rate. Assume the individual chooses surgery. This iterative process continues varying the probabilities in the gamble until the two options (living with anxiety vs taking a risk of curative surgery are equally desirable, that is, the respondent is indifferent between the gamble and living with anxiety). We make the assumption that the utility of perfect health is 1 and death is 0. For example, if this point were a surgical risk of 1%, the utility of living with anxiety about the adrenal mass is 0.99 and the disutility, defined as 1-utility would be 0.01. Utilities of other health states can be determined in a similar fashion. The time trade-off method is another way of determining utility. The utility may be described as a number between 0 and 1. For example, a subject may consider 10 years with pain from diabetic neuropathy equivalent to 5 years in perfect health. The utility of diabetic neuropathy would be 0.5. Alternatively, the utility may be expressed in quality-adjusted life years (QALYs). To determine the number of QALYs associated with an outcome, the time horizon for the outcome state is specified. Often, this is a patient's life expectancy in a particular outcome state. The number of years in full health that the subject sees as equivalent to the specified time with that outcome is the corresponding number of QALYs.

Determine the Probability of Each Chance Event

Once the decision tree structure is formed, the probability of each chance event may be determined. In general, this is best performed by taking a systematic review of published, peer-reviewed literature. However, this approach assumes that probabilities derived from a period in the past accurately reflect the probabilities in the future. Moreover, it assumes that the probabilities derived from other settings apply to the one in question. However, all these probabilities must reflect actual practice. For example, if decision analysis is being used to determine the best strategy for specific individual patients, the probabilities must be those of the site when the care is to be delivered. At times, not all of the information needed for the decision tree is available. This frequently occurs during analysis of a relatively new practice. If possible, primary data may be collected or secondary data may be analyzed. Expert opinion may be used in the absence of relevant data. At a given chance node, the sum of probabilities equals one.

Deciding on a Strategy: Averaging Out and Folding Back the Tree

The goal of decision analysis is to identify the strategy that leads to the most favorable expected outcome. To calculate the expected outcome, one starts at the outcome measures (typically the far right). Each group of branches, which start at a single chance node, is reduced to a single numerical value by multiplying each outcome utility by the probability associated with that outcome and adding all of the values. This is the process of folding back the decision tree (see Figure 3–10). This process is repeated until there is a single numerical value for each strategy at the initial decision node. At this point, the expected outcome for each strategy has been calculated. The strategy with the more favorable outcome is the preferred strategy.

Discounting Future Events

A greater value is placed on current events than future events. It is better to pay $100 in 10 years rather than pay $100 now. Similarly, if one is to have a disease, it is preferable to have a disease in the future than to have it today. The value of a future event then depends on how far in the future it occurs. Discounting refers to calculating the present value of an outcome that occurs in the future. The discount rate is the annual rate at which costs are discounted, which is usually the rate of interest that money would bring if it were invested.

Sensitivity Analysis

The initial analysis, or base case analysis, uses the best estimates for each parameter of the model. However, these estimates may be off the mark. The purpose of sensitivity analysis is to determine which estimates have the greatest influence on the model, that is, test the stability or robustness of a conclusion over a range of structural assumptions, value judgments, and estimates of probability. A plausible range of values is determined for each parameter of the model. This process determines which of the parameters drive the model's results. Parameters can be tested one at a time (one-way sensitivity analysis), two at a time (two-way sensitivity analysis), or even more (multi-way sensitivity analysis), though the latter become more difficult to interpret as the number of parameters tested is increased.

Cost-Effectiveness Analysis Using Decision Analysis

Cost-effectiveness analysis is the use of decision analysis to compare strategies in terms of their cost per unit of output. Output is an outcome such as years of life, utility, or cases of disease prevented. Cost-effectiveness ratios are interpreted by comparing them to ratios for other strategies. An incremental cost-effectiveness ratio indicates how much additional money needs to be spent for a better, but more expensive strategy to generate one additional unit of outcome.

Of practical importance, there usually is a limit to the amount of money a policy-maker is willing to spend to gain one QALY;

TABLE 3–2 Users' guide to decision analyses.

I. Are the results of the study valid?
   Were all important strategies and outcomes included?
   Was an explicit and sensible process used to identify, select, and combine the evidence into probabilities?
   Were the utilities obtained in an explicit and sensible way from credible sources?
   Was the potential impact of any uncertainty in the evidence determined?

II. What are the results?
   In the baseline analysis, does one strategy result in a clinically important gain for patients? If not, is the result a "toss-up"?
   How strong is the evidence used in the analysis?
   Could the uncertainty in the evidence change the result?

III. Will the results help me in caring for my patients?
   Do the probability estimates fit my patients' clinical features?
   Do the utilities reflect how my patients would value the outcomes of the decision?

Data from Richardson WS, Detsky AS for the Evidence-Based Medicine Working Group; based on the *Users' Guides to Evidence-Based Medicine* and reproduced with permission from *JAMA*. 1995 Apr 26;273:1292-1295 and Richardson WS, Detsky AS: Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients? Evidence Based Medicine Working Group, *JAMA*. 1995 May 24-31;273(20):1610-1613.

this is termed the willingness-to-pay threshold. Cost–utility analysis is a specific type of cost-effectiveness analysis that uses QALYs (or other measures of utility) as the effectiveness end-point. By convention, cost–utility analyses are often called cost-effectiveness analyses. However, not all cost-effectiveness studies use the cost–utility methodology. Because they use QALYs as an end-point, cost–utility analyses generate information that may be compared across disease states. It is important to note that although they are useful in providing a common metric among outcomes, a number of stringent assumptions must be made to compare cost-effectiveness analyses. Principles for assessing decision analysis have been developed by the EBM working group and are shown in Table 3–2. Representative decision analyses related to endocrine disorders are shown in Table 3–3.

OTHER ASPECTS OF CLINICAL EPIDEMIOLOGY

The methods of clinical epidemiology have been applied to risk assessment, prognostication, the assessment of studies of different designs, and many other issues in clinical medicine. Readers are referred to standard texts.

EVIDENCE-BASED ENDOCRINOLOGY

The definition of EBM has evolved over time. A 2005 definition states that EBM requires the integration of the best research evidence with clinical expertize and patient's unique values and circumstances. The best research evidence refers to valid and clinically relevant research, often from the basic sciences of medicine, but especially from patient-centered clinical research. Clinical expertize refers to the ability to use physicians' clinical skills and past experience to rapidly identify each patient's unique health state and diagnosis, their individual risks and benefits of potential interventions, and their personal circumstances and expectations. Patient values consist of the unique preferences, concerns, and expectations each patient brings to a clinical encounter that must be integrated into clinical decisions if they are to serve the patient. Patient circumstances refer to the patient's individual clinical state and the clinical setting.

The five steps of EBM as defined in 1997 are discussed later. Entire textbooks are devoted to the principles of EBM. In this chapter, we will briefly review these five steps and focus on how EBM views evidence and review some of the limitations.

Step One: Translation of the Clinical Problem into Answerable Questions

One of the fundamental skills required for practicing EBM is the asking of well-built clinical questions. To benefit patients and clinicians, such questions need to be both directly relevant to patients' problems and phrased in ways that direct a literature search to relevant and precise answers. In practice, well-built clinical questions usually contain four elements: patient, intervention, comparison, and outcome. The patient refers to the individual to whom the evidence will be applied (eg, a hypertensive patient with diabetes). (The evidence-based approach can also be applied to a group to inform policy.) The intervention is the diagnostic test or therapy being considered for application to the particular patient (or group) (eg, tight control of blood pressure). The comparison is that intervention to which a new intervention being considered is to be compared (eg, less tight control of blood pressure). The outcome is the end-point of interest to the physician or patient (eg, stroke, myocardial infarction, or death from diabetes). Background questions relate to general information or basic facts about a disease, and this information can usually be found in reliable textbooks, whether published in book form or on-line. However, more patient-specific clinical questions, especially those that require the most up-to-date information about diagnosis, prognosis, or treatment, have been termed foreground questions and require a different approach to find relevant evidence.

Step Two: Finding the Best Evidence

The finding of best research evidence from the medical literature is a critical feature of the EBM approach and is an iterative process that involves medical informatics. Given the enormous breadth and depth of the medical literature, with more than two million articles published each year in about 20,000 biomedical journals (with more being established each year), efficient strategies must be used to take advantage of the best original journal articles, reviews and synopses of primary studies, and practice guidelines, along with textbooks, both traditional and innovative. Among the useful sources of information are PubMed, which allows sophisticated search strategies of the MEDLINE database. A variety of tools have been developed to facilitate this process. However, these are beyond the scope of this chapter, and readers are referred to references on the subject. Certain journals are specifically devoted

TABLE 3-3 Representative decision analyses in endocrinology.

Study	Clinical Problem and Strategies	Time Horizon	Outcome Measures	Results
Diabetes				
DCCT	Conventional vs intensive insulin therapy in approximately 120,000 persons with IDDM in the United States who meet DCCT inclusion criteria	Lifetime	Years free from diabetic complications Cost per life-year	Intensive insulin therapy results in a gain of 920,000 y of sight, 691,000 y free from ESRD, 678,000 y free from lower extremity amputation, and 611,000 y of life at a cost of $4 billion. Cost per life-year gained is $28,661.
Hoerger	Screening for type 2 DM in all people vs only screening those with hypertension	Lifetime	Cost per QALY	Screening in people with hypertension is more cost-effective. Targeted screening for people age 55-75 is most cost-effective.
Gillies	Different strategies for screening and prevention of type 2 DM in adults	Lifetime	Cost per QALY	Screening for type 2 DM and impaired glucose tolerance, with appropriate intervention for those with impaired glucose tolerance, in an above average risk population aged 45 seems to be cost-effective.
Golan	To preserve kidney function in patients with DM type 2, strategies of treating all patients with ACE inhibitors vs screening for microalbuminuria vs screening for gross proteinuria	Lifetime	Cost per QALY	Screening for gross proteinuria has the highest cost and lowest benefit. Compared with microalbuminuria, treating all patients with an ACE-inhibitor was beneficial with an IC-ER of $7500 per QALY gained.
Gaede	Intensified vs conventional multifactorial intervention in type 2 DM	Lifetime	Cost per QALY	From a health-care perspective in Denmark, intensive therapy was more cost-effective than conventional treatment. Assuming that persons in both arms were treated in a primary care setting, intensive therapy was both cost- and life-saving.
ACCORD	Intensive vs standard glycemic control in older patients with type 2 DM	Lifetime	Survival, adverse cardiovascular events	Intensive glycemic control did not help older patients with type 2 DM who had the disease for longer duration and had higher risk for cardiovascular disease. In fact it may cause harm to these individuals
Fried	To preserve kidney function in patient with type 2 DM by combination treatment of angiotensin converting enzyme inhibitor and angiotensin II receptor blocker	Lifetime	Slowing progression of chronic kidney disease	Combination therapy was associated with increased risk of adverse effects with no clear benefit.
Thyroid				
Danese	In asymptomatic adults, screening for mild thyroid failure every 5 y starting at age 35 vs no screening	Lifetime	Cost per QALY	Screening was cost-effective with a cost per QALY gained of $9223 for women and $22,595 for men.
Vidal-Trécan	Four strategies to treat solitary toxic thyroid adenoma in a 40-y-old woman: (A) Primary radioactive iodine (B) Primary surgery after euthyroidism achieved by ATDs (C) ATDs followed by surgery or (D) ATDs followed by radioactive iodine C and D were used if severe reaction to ATDs occurred	Lifetime	Cost per QALY	Surgery was the most effective and least costly strategy. Primary radioactive iodine was more effective if surgical mortality exceeded 0.6%.
Nasuti	Evaluation of FNA by a cytopathologist with on-site processing vs standard processing at the University of Pennsylvania Medical Center	Short term	Cost	By avoiding nondiagnostic specimens, an estimated cost savings of $404,525 per year may be achieved with on-site FNA review.
McCartney	Analysis of different thyroid nodule biopsy guideline criteria	Short term	Utility (expected value)	As a general approach to 10-14 mm thyroid nodules, routine FNA appears to be the least desirable compared with observation alone or FNA of nodules with ultrasonographic risk factors.
Wang	Cost effectiveness analysis of papillary thyroid cancer surveillance after total thyroidectomy	3 y	Cost of surveillance for each recurrence	Cost to detect each recurrence of papillary thyroid cancer in a low risk patient was more than 6 and 7 times greater than the cost for intermediate and high risk patients. It was $147,819, $22,434 and $20,680 for low, intermediate and high risk cancer respectively.

(Continued)

TABLE 3-3 Representative decision analyses in endocrinology. (*Continued*)

Study	Clinical Problem and Strategies	Time Horizon	Outcome Measures	Results
Other				
King	Four strategies to manage inciden-tal pituitary microadenoma in an asymptomatic patient: (1) Expectant management (2) PRL screening (3) Screening for PRL, insulin-like growth factor I, and (4) MRI follow-up	Lifetime	Cost per QALY	PRL test may be the most cost-effective strategy. Compared to expectant management, the IC-ER for PRL was $1428. The IC-ER for the extended screening panel was $69,495. MRI follow-up was less effective and more expensive.
Sawka	Three strategies to evaluate pheo-chromocytoma in patients with refractory hypertension, suspi-cious symptoms, adrenal mass, or history of pheochromocytoma: (A) Fractionated plasma metaneph-rines with imaging if abnormal (B) 24-h urinary metanephrines or catecholamines with imaging if abnormal (C) Plasma metanephrines if mod-estly elevated, urine studies to decide on imaging	Short term	Cost per pheochro-mocytoma detected	Strategy C is least costly and has reasonable sensitivity in patients with moderate pretest probability for pheochromocytoma.
Col	For menopausal symptom relief in healthy, white, 50-y-old women with intact uteri, use of hormone therapy vs no hormone therapy	2 y	Survival QALE	Hormone therapy is associated with lower survival but gains in QALE. Benefits depend on severity of menopausal symp-toms and CVD risk.
Smith	For 60-y-old men with erectile dys-function, sildenafil vs no drug therapy	Lifetime	Cost per QALY gained	From the societal perspective, cost per QALY gained with sildenafil is less than $50,000 if treatment-related morbid-ity is less than 0.55% per year, treatment success rate is greater than 40.2%, or cost of sildenafil is less than $244 per month.
Ollendorf	Comparing effectiveness and value of various treatment options for obesity. Comparing four types of bariatric surgery (biliopancreatic diversion, laproscopic adjustable gastric banding, Roux-en-Y gastric bypass, vertical sleeve gastrec-tomy with conventional weight loss management	10 y	Cost per QALY gained	Among patients with BMI 30.0 to 34.9 the cost–effectiveness of bariatric surgery compared to conventional weight loss management was $40,000 to $60000 per QALY.

DCCT, Diabetes Control and Complications Trial; IDDM, insulin-dependent diabetes mellitus; ESRD, end-stage renal disease; DM, diabetes mellitus; QALY, quality-adjusted life year; IC-ER, incremental cost-effectiveness ratio; ACE, angiotensin-converting enzyme; CVD, cardiovascular disease; ATDs, antithyroid drugs; FNA, fine-needle aspiration; PRL, prolactin; MRI, magnetic resonance imaging; QALE, quality-adjusted life expectancy.

- DCCT. Lifetime benefits and costs of intensive therapy as practiced in the diabetes control and complications trial The Diabetes Control and Complications Trial Research Group. *JAMA*. 1996;276(17):1409-1415.
- Hoerger TJ, Harris R, Hicks KA, Donahue K, Sorensen S, Engelgau M. Screening for type 2 diabetes mellitus: a cost-effectiveness analysis. *Ann Intern Med*. 2004;140(9):689-699.
- Gillies CL, Lambert PC, Abrams KR, et al. Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. *BMJ*. 2008;336:1180-1185.
- Golan L, Birkmeyer JD, Welch HG. The cost-effectiveness of treating all patients with type 2 diabetes with angiotensin-converting enzyme inhibitors. *Ann Intern Med*. 1999;131(9):660-667.
- Gaede P, Valentine WJ, Palmer AJ, et al. Cost-effectiveness of intensified versus conventional multifactorial intervention in type 2 diabetes. *Diabetes Care*. 2008;31(8):1510-1515.
- The Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes mellitus. *N Engl J Med*. 2008;358:2545-2559.
- Fried LF, Emanuele N, Zhang JH, Brophy M, Conner TA, Duckworth W, Leehey DJ, McCullough PA, O'Connor T, Palevsky PM, Reilly RF, Seliger SL, Warren SR, Watnick S, Peduzzi P, Guarino P; VA NEPHRON-D Investigators. Combined angiotensin inhibition for the treatment of diabetic nephropathy. *N Engl J Med*. 2013 Nov 14;369(20):1892-903.
- Danese MD, Powe NR, Sawin CT, Ladenson PW. Screening for mild thyroid failure at the periodic health examination: a decision and cost-effectiveness analysis. *JAMA*. 1996;276(4):285-292.
- Vidal-Trecan GM, Stahl JE, Durand-Zaleski I. Managing toxic thyroid adenoma: a cost-effectiveness analysis. *Eur J Endocrinol*. 2002;146(3):283-294.
- Nasuti JF, Gupta PK, Baloch ZW. Diagnostic value and cost-effectiveness of on-site evaluation of fine-needle aspiration specimens: review of 5,688 cases. *Diagn Cytopathol*. 2002;27(1):1-4.
- McCartney CR, Stukerborg GJ. Decision analysis of discordant thyroid nodule biopsy guideline criteria. *J Clin Endocrinol Metab*. 2008;93(8):3037-3044.
- Wang LY, Roman BR, Migliacci JC, Palmer FL, Tuttle RM, Shaha AR, Shah JP, Patel SG, Ganly I, Cost-effectiveness analysis of papillary thyroid cancer surveillance. *Cancer*. 2015 Dec 1;121(23):4132-40.
- King JT, Jr, Justice AC, Aron DC. Management of incidental pituitary microadenomas: a cost-effectiveness analysis. *J Clin Endocrinol Metab*. 1997;82(11):3625-3632.
- Sawka AM, Gafni A, Thabane L, Young WF, Jr. The economic implications of three biochemical screening algorithms for pheochromocytoma. *J Clin Endocrinol Metab*. 2004;89(6):2859-2866.
- Col NF, Weber G, Stiggelbout A, Chuo J, D'Agostino R, Corso P. Short-term menopausal hormone therapy for symptom relief: an updated decision model. *Arch Intern Med*. 2004;164(15):1634-1640.
- Smith KJ, Roberts MS. The cost-effectiveness of sildenafil. *Ann Intern Med*. 2000;132(12):933-937.
- Ollendorf DA, Cameron CG, Pearson SD. Effectivenss and value of treatment options for obesity – a report for the California Technology Assessment Forum. *JAMA Int Med*. 2106; 176:247-248.

to critically appraised articles (eg, ACP Journal Club, BMJ Evidence-Based Medicine). It is important to remember, however, that the presence of the label *evidence based* does not ensure that the information meets the highest standards. Systematic reviews are another useful source for evidence. In addition to the professional journals, such reviews are available from both governmental agencies (eg, Agency for Healthcare Research and Quality) and nongovernmental agencies (eg, the Cochrane Collaboration). Standards have been developed for reporting systematic reviews: preferred reporting items for systematic reviews and meta-analyses (PRISMA). Among the issues addressed specifically is that of publication bias and its potential effects on the results. Publication bias reflects, among other things, the tendency to not publish studies with negative findings.

Step Three: Appraising the Evidence for Its Validity and Usefulness

Critical appraisal is the process of systematically examining research evidence to assess its validity, results, and relevance before using it to inform a decision. There are three basic questions to be addressed in evaluating any kind of research: (1) Is it valid? (2) Is it important? and (3) Is it applicable to my patient? This involves assessment of the study's methods in order to assess the internal validity of the conclusions. Once criteria for internal validity are satisfied, then the importance of the finding can be assessed. It is critical that statistical significance not be equated with clinical significance. Assuming the finding is important, it is necessary to consider the study's relevance to the patient (ie, the external validity of the study). A series of guides has been developed to help the clinician to critically appraise different types of studies. An example is shown in Table 3–4. Underlying this appraisal process is the EBM model of the hierarchy of evidence.

A. Hierarchies of evidence   Any empirical observation about the apparent relation between events constitutes evidence. This includes the thoughtful observations of an experienced clinician, the observations derived from physiologic experiments, and the results of large, randomized clinical trials. Methodological safeguards to protect against systematic error (bias) are inherent in experimental design. In this way, experiments differ from clinical observations. Different study designs lend themselves to different safeguards, and these safeguards vary in their efficacy both in theory and in practice.

The EBM approach uses a hierarchy of study designs based on their inherent safeguards. A hierarchy is shown in Table 3–5. In this hierarchy, randomized controlled trials (RCTs) are given the highest weight, followed by quasi-experimental studies (ie, no-RCTs), followed by other study designs, and finally opinions of authorities and unsystematic clinical observations. Evidence from the top of the hierarchy should have greater impact in clinical decision-making than observations from the lower levels. There is disagreement about what the hierarchy should be. Some have an N of 1 randomized trial (the equivalent of an RCT with one subject) at the top of the hierarchy, whereas others place meta-analyses of randomized trials at the top. However, the term *evidence* is

---

TABLE 3–4 Users' guides for an article about therapy.

I. Are the results of the study valid?
Primary Guides
Was the assignment of patients to treatments randomized?
Were all patients who entered the trial properly accounted for and attributed at its conclusion?
Was follow-up complete?
Were patients analyzed in the groups to which they were randomized?
Secondary Guides
Were patients, health workers, and study personnel "blind" to treatment?
Were the groups similar at the start of the trial?
Aside from the experimental intervention, were the groups treated equally?
II. What were the results?
How large was the treatment effect?
How precise was the estimate of the treatment effect?
III. Will the results help me in caring for my patients?
Can the results be applied to my patient care?
Were all clinically important outcomes considered?
Are the likely treatment benefits worth the potential harms and costs?

Data from Guyatt GH, Sackett D, Cook DJ, for the Evidence Based Medicine Working Group. Based on the *Users' Guides to Evidence-Based Medicine* and reproduced with permission from *JAMA.* 1993;270:2598-2601 and Richardson WS, Detsky AS: Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients? Evidence Based Medicine Working Group, *JAMA.* 1995 May 24-31;273(20):1610-1613.

---

not synonymous with evidence from RCTs (or whatever study design occupies at the preferred level of the hierarchy). Rather, it is preferred to use that sort of evidence when it is available. Evidence from other sources is still evidence and may be more than sufficient. For example, the introduction of thyroid replacement therapy for hypothyroidism or insulin therapy for type 1 diabetes has never been subjected to RCTs.

While recognizing the central role of evaluating the quality of the evidence in decision-making, the U.S. Preventive Services Task Force (USPSTF) has developed a different approach to address some of the shortcomings of the EBM approach. The USPSTF adopted three major changes to the process. First, a rating of internal validity was added to the study criterion for judging individual studies. A well-designed cohort or case-control study could be more compelling than an inadequately powered or poorly

TABLE 3–5 A hierarchy of evidence.

1. Evidence obtained from at least one properly designed randomized controlled trial.
2a. Evidence obtained from well-designed controlled trials without randomization.
2b. Evidence obtained from well-designed cohort or case-control analytical studies, preferably from more than one center or research group.
2c. Evidence from comparisons between times or places with or without the intervention. Dramatic results in uncontrolled experiments could be included in this section.
3. Opinions of respected authorities, based on clinical experience, descriptive studies, or reports on expert committees.

conducted RCT. Second, evidence was explicitly assessed at three different levels—the level of the individual study, both in terms of internal and external validity; the body of evidence supporting a given linkage in the analytic framework; and, because of the focus of the USPSTF, the level of evidence for the entire preventive service. Third, the magnitude of effect was separated from the assessment of quality. Both benefit and harm are considered in assessing the magnitude of the effect(s). Overall, the USPSTF grades the quality of the overall evidence for a service on a three-point scale (good, fair, poor) and grades its recommendations according to one of five classifications (A, B, C, D, I) reflecting the strength of evidence and magnitude of net benefit (benefits minus harms) (Table 3–6). This approach of assessing both the quality of evidence and strength of recommendations has been expanded to include practice guidelines: grading of recommendations, assessment, development, and evaluation (GRADE) system. Although a discussion of practice guidelines is beyond the scope of this chapter, it is important to recognize that they vary widely in quality, strength of evidence, and bias. Before a practitioner bases a decision on a practice guideline, the basis of that guideline must be clear.

TABLE 3-6 The USPSTF grading system for recommendations reflecting the strength of evidence and magnitude of net benefit (benefits minus harms).

Grade	Definition	Suggestions for Practice
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.	Offer or provide this service.
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.	Offer or provide this service.
C	The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small.	Offer or provide this service for selected patients depending on individual circumstances.
D	The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.	Discourage the use of this service.
I Statement	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.	Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.
Levels of Certainty Regarding Net Benefit		
Level of Certainty[*]	Description	
High	The available evidence usually includes consistent results from well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on health outcomes. This conclusion is therefore unlikely to be strongly affected by the results of future studies.	
Moderate	The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by such factors as: • The number, size, or quality of individual studies. • Inconsistency of findings across individual studies. • Limited generalizability of findings to routine primary care practice. • Lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion.	
Low	The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of: • The limited number or size of studies. • Important flaws in study design or methods. • Inconsistency of findings across individual studies. • Gaps in the chain of evidence. • Findings not generalizable to routine primary care practice. • Lack of information on important health outcomes. More information may allow estimation of effects on health outcomes.	

[*]The USPSTF defines certainty as "likelihood that the USPSTF assessment of the net benefit of a preventive service is correct." The net benefit is defined as benefit minus harm of the preventive service as implemented in a general, primary care population. The USPSTF assigns a certainty level based on the nature of the overall evidence available to assess the net benefit of a preventive service.

Reproduced with permission from U.S. Preventive Services Task Force Ratings. Strength of recommendations and quality of evidence. http://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions.

B. Other evidence-related issues   Notwithstanding these improvements, there are four other major issues concerning the evidence. First, it is important to recognize that virtually all of this evidence is essentially quantitative in nature. Thus, the EBM definition of evidence is relatively narrow and excludes information important to clinicians; many qualitative factors are involved in clinical decision-making concerning individual patients. Second, there are problems in the evidence in terms of the quality and scope of the data (eg, the soft data that clinicians usually use); the scope of the topics (average patients, gray zones of practice, lack of RCTs); sources of authority, which however explicit, require judgment; and potential abuses (eg, key findings could be ignored). Third, there is the issue of applying results in populations (ie, average results) to individuals. Fourth, there is the dynamic nature of evidence.

Evidence to support practice of medicine is always evolving with ongoing research in medical field and in that sense evidence would be dynamic. Whenever there is a change in evidence regarding an intervention, the best practice of medicine should incorporate that evidence into how clinical intervention is carried out. At times, new evidence may refute prior evidence, requiring a reversal in practice. In this case, health-care provider would need to "de-adopt" or "de-implement" a previously carried out medical practice. A good example is Action to Control Cardiovascular Risk in Diabetes (ACCORD) study failed to demonstrate clear benefit in survival for older patients with type 2 diabetes who were treated to achieve intensive control of blood glucose. Subsequent to the results of this study, both health-care providers and patients have to reset the target for blood glucose control to a higher range or less intensive control of blood glucose in this patient population.

Steps Four and Five: Applying the Results in Practice and Evaluating Performance

Knowledge of the best available evidence of treatment is not enough to recommend a particular course of action. Underlying this issue is heterogeneity of treatment effects. **Treatment effect heterogeneity** is the term given to the phenomenon in which the same treatment produces different results in different patients. For the average effect observed in a clinical trial to occur with certainty in one patient, all patients in the trial must have had that average response. Yet average response may also reflect a large response in some and small or no response or even harm in others. Part of clinical decision-making involves addressing this phenomenon to individualize therapy in the most effective way. To do so, four major factors about the patient must be taken into account: (1) baseline probability of incurring a disease-related adverse event (risk without treatment or susceptibility/prognosis), (2) responsiveness to the treatment, (3) vulnerability to the adverse side-effects of the treatment, and (4) utilities for different outcomes. When lacking good data on a patient's individual level of risk, responsiveness, and vulnerability, the average treatment effect as reported in good clinical trials provides the most reasonable guide for decision-making. However, it is important that physicians evaluate the results in their own patients so that the course can be modified when necessary.

Developments That May Affect the EBM Approach

Two developments are beginning to have an impact on EBM: "Big Data" and "Personalized Medicine." The global storage capacity for data has increased greatly and continues to do so. There are now data sets so large that the usual approaches do not work very well. Predictive analytics, a burgeoning field can identify new correlations to spot trends in a wide variety of fields including medicine. In contrast to the EBM approach which emphasizes the explicit control of biases in data collection and analysis, the "big data" approach generally involves the maxim "more data are better data." However, this approach can complement typical EBM methods. Another development is that of personalized medicine, also referred to as "precision medicine" in which patient genetic information (sometimes arrived at via a big data approach) is used to inform patient-specific treatment. This patient-specific information can complement evidence from the EBM hierarchy, which emphasizes what works on average. However, some things are unlikely to change. Practicing and perfecting the art of medicine demands recognition that uncertainty permeates all clinical decisions. How clinicians make decisions, whether diagnostic or therapeutic, by combining the art and the science is a complicated matter replete with unknowns. There is a variety of modes of clinical problem-solving ranging from automatic decisions in which a clinician makes a decision without consciously considering alternatives or even specific features of the problem to rational decision-making using simplified strategies to take into account probabilities and values of outcomes to formal decision analysis. The potential deficiencies of automatic decisions (nondecisions one might say) are obvious. Formal decision analyses, even if accepted as gold standards (which they are not), are beyond practical use for the vast majority of clinicians, so that we are left with simplified processes to combine likelihoods and values. However, even in striving to reduce uncertainty and coming up with the best course of action for the patient, it is critical to recognize that uncertainty will always be with us. In using a quantitative approach, we can only become more certain about the probability of the outcome. We cannot guarantee the desired outcome.

REFERENCES

General

Greenberg RS, Daniels SR, Flanders WD, et al. *Medical Epidemiology*. New York, NY: McGraw-Hill Medical; 2005.

Grobbee DE, Hoes AW. *Clinical Epidemiology: Principles, Methods, and Applications for Clinical Research*. London: Jones & Bartlett Publishing; 2009.

Guyatt G, Rennie D, Meade M, Cook D. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice*. 2nd ed (*JAMA* & Archives Journals). New York, NY: McGraw-Hill; 2008.

Janicek M. *The Foundations of Evidence-Based Medicine*. New York, NY: Parthenon Publishing Group; 2003.

Jenicek M, Hitchcock DL. *Evidence-Based Practice: Logic and Critical Thinking in Medicine*. Chicago, IL: AMA Press; 2005.

Mayer D. *Essential Evidence-Based Medicine*. New York, NY: Cambridge Medicine; 2009.

Montori VM, ed. *Evidence-Based Endocrinology*. Tolowa: Humana Press Inco; 2006.

Muennig P. *Cost-Effectiveness in Health: A Practical Approach*. New York, NY: John Wiley and Sons; 2009.

Strauss SE, Richardson WS, Glasziou P, Haynes, RB. *Evidence-Based Medicine*. 3rd ed. Churchill Livingstone; 2005.

Diagnostic Testing and Treatment Decisions (Clinical Decision-Making and Decision Analysis)

Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press; 2015.

Montori VM, Wyer P, Newman TB, et al. Evidence-Based Medicine Teaching Tips Working Group: Tips for learners of evidence-based medicine. 5. The effect of spectrum of disease on the performance of diagnostic tests. *CMAJ*. 2005;173:385. [PMID: 16103513]

Myriam HMG, Weinstein MC, Wittenberg E, et al. *Decision Making in Health and Medicine: Integrating Evidence and Values*. Cambridge UK: Cambridge University Press; 2014.

Schwartz A, Bergus G. *Medical Decision Making: A Physician's Guide*. New York, NY: Cambridge University Press; 2008.

Evidence-Based Medicine: General

Berwick D. Broadening the view of evidence-based medicine. *Qual Saf Health Care*. 2005;14:315. [PMID: 16195561]

Daly J. *Evidence-Based Medicine and the Search for a Science of Clinical Care*. California: University of California Press/Milbank Memorial Fund; 2005.

Greenhalgh T. Integrating qualitative research into evidence based practice. *Endocrinol Metab Clin North Am*. 2002;31:583. [PMID: 12227121]

Translation of the Clinical Problem into Answerable Questions

Bergus GR, Randall CS, Sinift SD. Does the structure of clinical questions affect the outcome of curbside consultations with specialty colleagues? *Arch Fam Med*. 2000;9:541. [PMID: 10862217]

Villanueva EV, Burrows EA, Fennessy PA, Rajendran M, Anderson JN. Improving question formulation for use in evidence appraisal in a tertiary care setting: a randomised controlled trial. *BMC Med Inf Decis Making*. 2001;I:4. [PMID: 11716797]

Finding the Best Evidence

Haynes RB, Walker-Dilks CJ. Evidence-based endocrinology: finding the current best research evidence for practice in endocrinology. *Endocrinol Metab Clin North Am*. 2002;31:527. [PMID: 12227117]

Richter B, Clar C. Systematic reviews in endocrinology: the Cochrane Metabolism and Endocrine Disorders Review Group. *Endocrinol Metab Clin North Am*. 2002;31:613. [PMID: 12227123]

Critical Appraisal and Applying the Results

Glasziou P, Vandenbroucke J, Chalmers I. Assessing the quality of research. *BMJ*. 2004;328:39. [PMID: 14703546]

Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *ACP J Club*. 2002;136:A11. [PMID: 11874303]

Helfand M, Morton SC, Guallar E, Mulrow C. A guide to this supplement. Challenges of summarizing better information for better health: the evidence-based practice center experience. *Ann Intern Med*. 2005;142(suppl):1033.

Ioannidis JP. How many contemporary medical practices are worse than doing nothing or doing less? *Mayo Clin Proc*. 2013;88:779. [PMID: 23871231].

Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001;286:821. [PMID: 11497536]

Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151:W-65. www.annals.org. [PMID: 1000100]

Montori VM, Jaeschke R, Schünemann HJ, et al. Users' guide to detecting misleading claims in clinical research reports. *BMJ*. 2004;329:1093. [PMID: 15528623]

Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA*. 2014;312(2):171. [PMID: 25005654]

U.S. Preventive Services Task Force Ratings. Strength of Recommendations and Quality of Evidence. *Guide to Clinical Preventive Services*. 3rd ed. Periodic Updates, 2000-2003. Rockville, MD: Agency for Healthcare Research and Quality. http://www.ahrq.gov/clinic/3rduspstf/ratings.htm.

van Dam HA, van der Horst F, van den Borne B, et al. Provider–patient interaction in diabetes care: effects on patient self-care and outcomes. A systematic review. *Patient Educ Couns*. 2003;51:17. [PMID: 12915276]

Wang MT, Gamble G, Grey A. Responses of specialist societies to evidence for reversal of practice. *JAMA Intern Med*. 2015;175:845. [PMID: 25774960].

Whitney SN, McGuire AL, McCullough LB. A typology of shared decision making, informed consent, and simple consent. *Ann Intern Med*. 2003;140:54. [PMID: 14706973]

Practice Guidelines

Aron D, Pogach L. Transparency standards for diabetes performance measures. *JAMA*. 2009;301:210. [PMID: 19141769]

Boyd CM, Darer J, Boult C, et al. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications of pay for performance. *JAMA*. 2005;294:716. [PMID: 16091574]

Brozek JL, Akl EA, Jaeschke R, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. *Allergy*. 2009;64:669. [PMID: 19210357]

Guyatt GH, Henry D, Hill S, et al. Rating quality of evidence and strength of recommendations: going from evidence to recommendations. *BMJ*. 2008;336:1049. [PMID: 18467413]

Guyatt GH, Oxman AD, Kunz R, et al. Rating quality of evidence and strength of recommendations: what is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008;336:995. [PMID: 18456631]

Guyatt GH, Oxman AD, Kunz R, et al. Rating quality of evidence and strength of recommendations: incorporating considerations of resources use into grading recommendations. *BMJ*. 2008;336:1170. [PMID: 18497416]

Guyatt GH, Oxman AD, Vist G, et al. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924. [PMID: 18436948]

Jaeschke R, Guyatt GH, Dellinger P, et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ*. 2008;337:a744. [PMID: 18669566]

McGoey L. Sequestered evidence and the distortion of clinical practice guidelines. *Persp Biol Med*. 2009;52:203. [PMID: 19395820]

Qaseem A, Vijan S, Snow V, Cross JT, Weiss KB, Owens DK. Clinical Efficacy Assessment Subcommittee of the American College of Physicians. Glycemic control and type 2 diabetes mellitus: the optimal hemoglobin A1c targets: a guidance statement from the American College of Physicians. *Ann Intern Med*. 2007;147:417. [PMID: 17876024]

Schünemann HJ, Oxman AD, Brozek J. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336:1106. [PMID: 18483053]

Shaneyfelt TM, Centor RM. Reassessment of clinical practice guidelines: go gently into that good night. *JAMA*. 2009;301:868. [PMID: 19244197]

Internet Sites Related to Evidence-Based Medicine

Center for Health Evidence. http://www.cebm.net/

Center for Evidence Based Medicine. http://www.ahrq.gov/clinic/epc/

Evidence-Based Medicine Web Resources. University of Washington Health Sciences Library. http://healthlinks.washington.edu/howto/ebmtable.html

Health Information Resource Unit, McMaster University. http://hiru.mcmaster.ca/

New Developments: Big Data and Personalized Medicine

Bray MS, Loos RJ, McCaffery JM, et al. NIH working group report-using genomic information to guide weight management: from universal to precision treatment. *Obesity*. 2016;24:14. [PMID: 26692578]

Cichosz SL, Johansen MD, Hejlesen O. Toward big data analytics: review of predictive models in management of diabetes and its complications. *J Diabetes Sci Technol*. 2015;10:27. [PMID: 26468133]

Kleinberger JW, Pollin TI. Personalized medicine in diabetes mellitus: current opportunities and future prospects. *Ann N Y Acad Sci*. 2015;1346:45. [PMID: 25907167]

Mega JL, Sabatine MS, Antman EM. Population and personalized medicine in the modern era. *JAMA*. 2014;312:1969. [PMID: 25399267]

Pearson E. Personalized medicine in diabetes: the role of 'omics' and biomarkers. *Diabet Med*. 2016;33:712. [PMID: 26802434]

Sim I. Two ways of knowing: big data and evidence-based medicine. *Ann Intern Med*. 2016;164:562. doi: 10.7326/M15-2970. (Epub 2016 Jan 26) [PMID: 26809201]

Tu JV, Chu A, Donovan LR, et al. The Cardiovascular Health in Ambulatory Care Research Team (CANHEART): using big data to measure and improve cardiovascular health and healthcare services. *Circ Cardiovasc Qual Outcomes*. 2015;8:204. [PMID: 25648464]

This page intentionally left blank

Hypothalamus and Pituitary Gland

Bradley R. Javorsky, MD, David C. Aron, MD, MS,
James W. Findling, MD, and J. Blake Tyrrell, MD

ACTH	Adrenocorticotropic hormone (corticotropin)		**IRMA**	Immunoradiometric assay
ADH	Antidiuretic hormone (vasopressin)		**KAL1**	Kallmann syndrome 1
CLIP	Corticotropin-like intermediate lobe peptide		**LH**	Luteinizing hormone
CRH	Corticotropin-releasing hormone		**β-LPH**	β-Lipotropin
CRHBP	Corticotropin-releasing hormone-binding protein		**Met-Enk**	Methionine-enkephalin
FGF8	Fibroblast growth factor 8		**MEN**	Multiple endocrine neoplasia
FGFR1	Fibroblast growth factor receptor 1		**MSH**	Melanocyte-stimulating hormone
FSH	Follicle-stimulating hormone		**PIH**	Prolactin-inhibiting hormone (dopamine)
GABA	Gamma-aminobutyric acid		**Pit-1**	Pituitary-specific positive transcription factor 1
GH	Growth hormone (somatotropin)		**POMC**	Pro-opiomelanocortin
GHBP	Growth hormone–binding protein		**PROK2**	Prokineticin 2
GHIH	Growth hormone–inhibiting hormone (somatostatin)		**PROKR2**	Prokineticin receptor 2
GHRH	Growth hormone–releasing hormone		**Prop-1**	Prophet of Pit-1
GHS-R	Growth hormone secretagogue receptor		**PRL**	Prolactin
GnRH	Gonadotropin-releasing hormone		*PTTG*	Pituitary tumor transforming gene
hCG	Human chorionic gonadotropin		**SHBG**	Sex hormone–binding globulin
hMG	Human menopausal gonadotropin		**SIADH**	Syndrome of inappropriate secretion of antidiuretic hormone
hPL	Human placental lactogen		**TRH**	Thyrotropin-releasing hormone
ICMA	Immunochemiluminescent assay		**TSH**	Thyroid-stimulating hormone (thyrotropin)
IGF	Insulin-like growth factor		**VIP**	Vasoactive intestinal peptide

The hypothalamus and pituitary gland form a unit that exerts control over the function of several endocrine glands—thyroid, adrenals, and gonads—as well as a wide range of physiologic activities. This unit is highly conserved across vertebrate species and constitutes a paradigm of neuroendocrinology—brain–endocrine interactions. The actions and interactions of the endocrine and nervous systems, whereby the nervous system regulates the endocrine system and endocrine activity modulates the activity of the central nervous system, constitute the major regulatory mechanisms for virtually all physiologic activities. These neuroendocrine interactions are also important in pathogenesis. This chapter reviews the normal functions of the pituitary gland, the neuroendocrine control mechanisms of the hypothalamus, and the disorders of those mechanisms.

Nerve cells and endocrine cells, which are both involved in cell-to-cell communication, share certain characteristic features—secretion of chemical messengers (neurotransmitters or hormones) and electrical activity. A single chemical messenger—peptide or amine—can be secreted by neurons as a neurotransmitter or neural hormone and by endocrine cells as a classic hormone. Examples of such multifunctional chemical messengers are shown in Table 4–1. The cell-to-cell communication may occur by four mechanisms:

TABLE 4–1 Neuroendocrine messengers: substances that function as neurotransmitters, neural hormones, and classic hormones.

	Neurotransmitter (Present in Nerve Endings)	Hormone Secreted by Neurons	Hormone Secreted by Endocrine Cells
Dopamine	+	+	+
Norepinephrine	+	+	+
Epinephrine	+		+
Somatostatin	+	+	+
Gonadotropin-releasing hormone (GnRH)	+	+	+
Thyrotropin-releasing hormone (TRH)	+	+	
Oxytocin	+	+	+
Vasopressin	+	+	+
Vasoactive intestinal peptide	+	+	
Cholecystokinin (CCK)	+		+
Glucagon	+		+
Enkephalins	+		+
Pro-opiomelanocortin derivatives	+		+
Other anterior pituitary hormones	+		+

(1) autocrine communication via messengers that diffuse in the interstitial fluid and act on the cells that secreted them, (2) neural communication via synaptic junctions, (3) paracrine communication via messengers that diffuse in the interstitial fluid to adjacent target cells (without entering the bloodstream), and (4) endocrine communication via circulating hormones (Figure 4–1). The two major mechanisms of neural regulation of endocrine function are direct innervation and neurosecretion (neural secretion of hormones). The adrenal medulla, kidney, parathyroid gland, and pancreatic islets are endocrine tissues that receive direct autonomic innervation (see Chapters 9, 10, and 11). An example of neurosecretory regulation is the hormonal secretion of certain hypothalamic nuclei into the portal hypophysial vessels, which regulate the hormone-secreting cells of the anterior lobe of the pituitary.

Another example of neurosecretory regulation is the posterior lobe of the pituitary gland, which is made up of the endings of neurons whose cell bodies reside in hypothalamic nuclei. These neurons secrete vasopressin and oxytocin into the general circulation.

Anatomy and Embryology

The anatomic relationships between the pituitary and the main nuclei of the hypothalamus are shown in Figure 4–2. The posterior lobe of the pituitary (neurohypophysis) is of neural origin, arising embryologically as an evagination of the ventral hypothalamus and the third ventricle. The neurohypophysis consists of the axons and nerve endings of neurons whose cell bodies reside in the supraoptic and paraventricular nuclei of the hypothalamus and

	Gap junctions	Synaptic	Paracrine	Endocrine
Message transmission	Directly from cell to cell	Across synaptic cleft	By diffusion in interstitial fluid	By circulating body fluids
Local or general	Local	Local	Locally diffuse	General
Specificity depends on	Anatomic location	Anatomic location and receptors	Receptors	Receptors

FIGURE 4–1 Intercellular communication by chemical mediators.
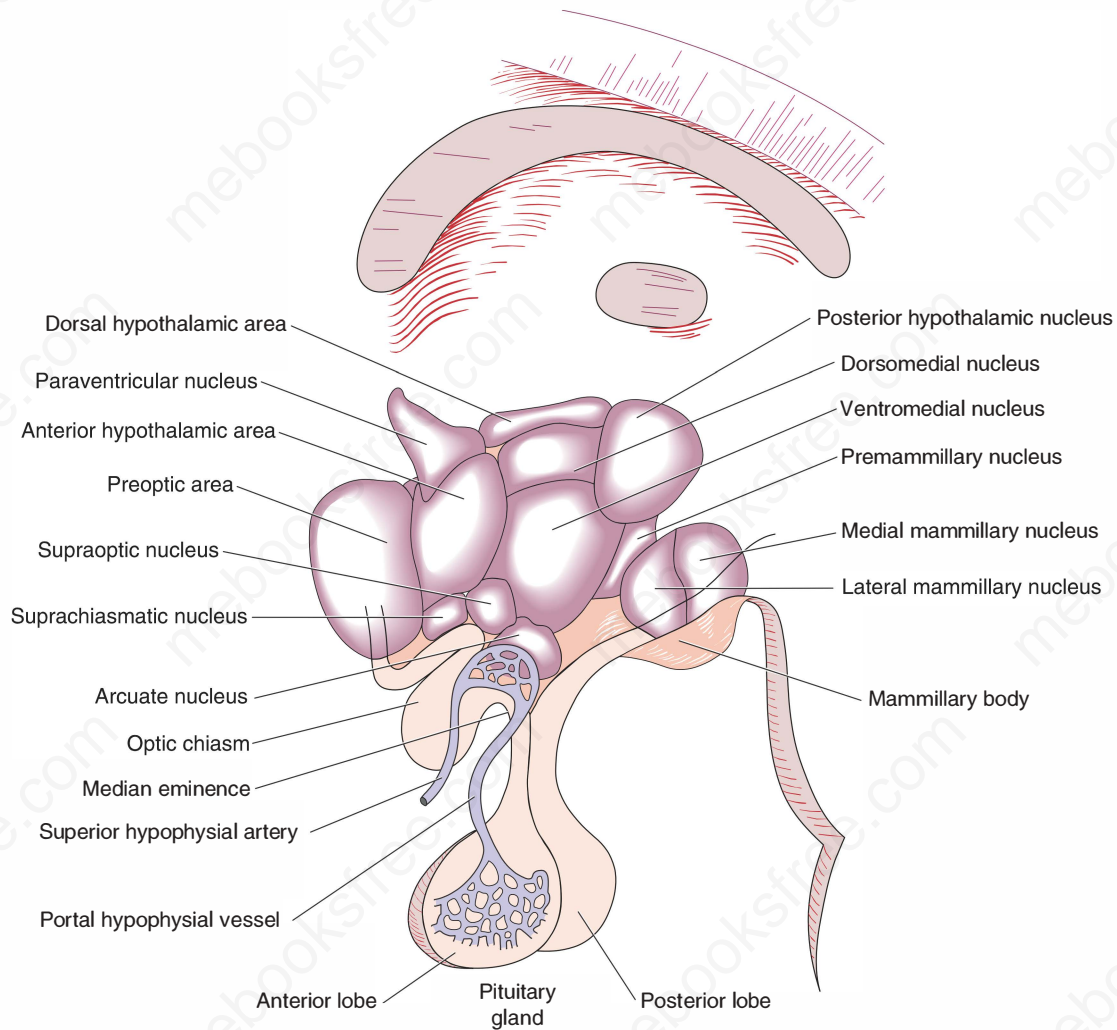
FIGURE 4–2 The human hypothalamus with a superimposed diagrammatic representation of the portal hypophysial vessels. (Reproduced with permission from Ganong WF. *Review of Medical Physiology*. 15th ed. New York: McGraw-Hill Education; 1993.)

supporting tissues. This hypothalamic-neurohypophysial nerve tract contains approximately 100,000 nerve fibers. Repeated swellings along the nerve fibers ranging in thickness from 1 to 50 μm constitute the nerve terminals.

The human fetal anterior pituitary anlage is initially recognizable at 4 to 5 weeks of gestation, and rapid cytologic differentiation leads to a mature hypothalamic-pituitary unit at 20 weeks. The anterior pituitary (adenohypophysis) originates from Rathke pouch, an ectodermal evagination of the oropharynx, and migrates to join the neurohypophysis. The portion of Rathke pouch in contact with the neurohypophysis develops less extensively and forms the intermediate lobe. This lobe remains intact in some species, but in humans its cells become interspersed with those of the anterior lobe and develop the capacity to synthesize and secrete pro-opiomelanocortin (POMC) and adrenocorticotropic hormone (ACTH). Remnants of Rathke pouch may persist at the boundary of the neurohypophysis, resulting in small colloid cysts. In addition, cells may persist in the lower portion of Rathke pouch beneath the sphenoid bone, the pharyngeal pituitary. These cells

have the potential to secrete hormones and have been reported to undergo adenomatous change.

The pituitary gland itself lies at the base of the skull in a portion of the sphenoid bone called the sella turcica (Turkish saddle). The anterior portion, the tuberculum sellae, is flanked by posterior projections of the sphenoid wings, the anterior clinoid processes. The dorsum sellae forms the posterior wall, and its upper corners project into the posterior clinoid processes. The gland is surrounded by dura, and the roof is formed by a reflection of the dura attached to the clinoid processes, the diaphragma sellae. The arachnoid membrane and, therefore, cerebrospinal fluid are prevented from entering the sella turcica by the diaphragma sellae. The pituitary stalk and its blood vessels pass through an opening in this diaphragm. The lateral walls of the gland are in direct apposition to the cavernous sinuses and separated from them by dural membranes. The optic chiasm lies 5 to 10 mm above the diaphragma sellae and anterior to the stalk (Figure 4–3).

The size of the pituitary gland, of which the anterior lobe constitutes two-thirds, varies considerably. It measures approximately
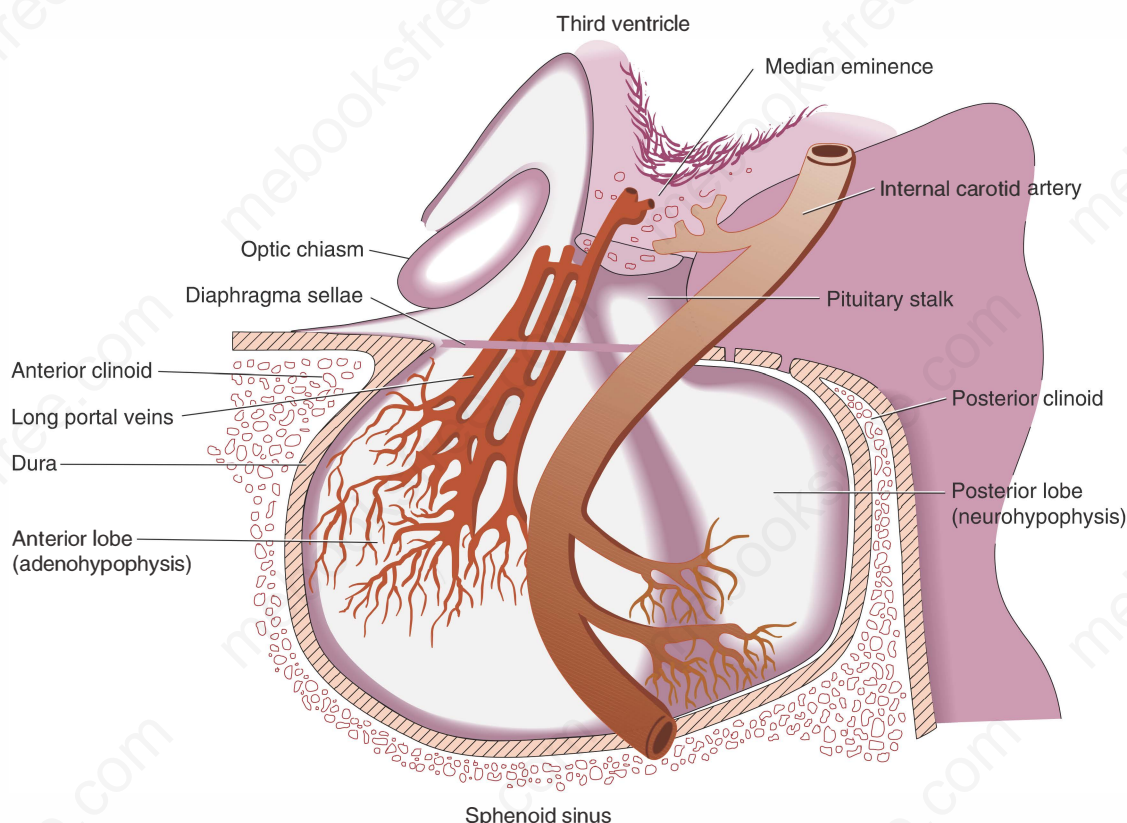
FIGURE 4–3 Anatomic relationships and blood supply of the pituitary gland. (Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism*. 3rd ed. New York: McGraw-Hill Education; 1995.)

15 × 10 × 6 mm and weighs 500 to 900 mg; it may double in size during pregnancy. The sella turcica tends to conform to the shape and size of the gland, and for that reason, there is considerable variability in its contour.

Blood Supply

The anterior pituitary is the most richly vascularized of all mammalian tissues, receiving 0.8 mL/g/min from a portal circulation connecting the median eminence of the hypothalamus and the anterior pituitary. Arterial blood is supplied from the internal carotid arteries via the superior, middle, and inferior hypophysial arteries. The superior hypophysial arteries form a capillary network in the median eminence of the hypothalamus that recombines in long portal veins draining down the pituitary stalk to the anterior lobe, where they break up into another capillary network and re-form into venous channels (see Figures 4–2 and 4–3). The pituitary stalk and the posterior pituitary are supplied directly from branches of the middle and inferior hypophysial arteries.

Venous drainage of the pituitary, the route through which anterior pituitary hormones reach the systemic circulation, is variable, but venous channels eventually drain via the cavernous sinus posteriorly into the superior and inferior petrosal sinuses to the jugular bulb and vein (Figure 4–4). The axons of the neurohypophysis terminate on capillaries that drain via the posterior lobe veins and the cavernous sinuses to the general circulation.

The hypophysial portal system of capillaries allows control of anterior pituitary function by the hypothalamic hypophysiotropic hormones secreted into the portal hypophysial vessels. This provides a short, direct connection to the anterior pituitary from the ventral hypothalamus and the median eminence (Figure 4–5). There may also be retrograde blood flow between the pituitary and hypothalamus, providing a possible means of direct feedback between pituitary hormones and their neuroendocrine control centers.

Pituitary Development and Histology

Anterior pituitary cells were originally classified as acidophils, basophils, and chromophobe cells based on staining with hematoxylin and eosin. Immunocytochemical and electron microscopic techniques now permit classification of cells by their specific secretory products: somatotrophs (growth hormone [GH]-secreting cells), lactotrophs (prolactin [PRL]-secreting cells), thyrotrophs (cells secreting thyroid-stimulating hormone [thyrotropin; TSH]), corticotrophs (cells-secreting ACTH [corticotropin] and related peptides), and gonadotrophs (luteinizing hormone [LH]– and follicle-stimulating hormone [FSH]–secreting cells). The development of the pituitary gland and the emergence of the distinct cell types from common primordial cells are controlled by a limited set of transcription factors, most notably Prop1 and Pit1 (Figure 4–6). The individual hormone-secreting
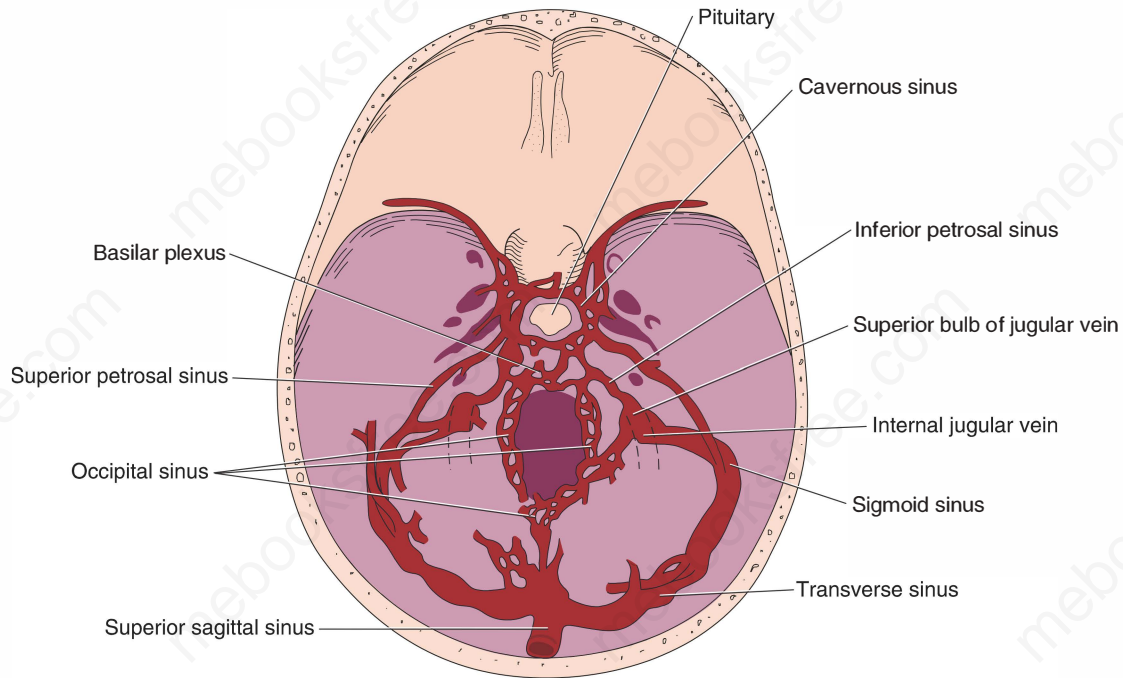
FIGURE 4–4   Venous drainage of the pituitary gland—the route by which adenohypophysial hormones reach the systemic circulation. (Reproduced with permission from Findling JW, Aron DC, Tyrrell JB, et al. Selective venous sampling for ACTH in Cushing's syndrome: differentiation between Cushing's disease and the ectopic ACTH syndrome. *Ann Intern Med.* 1981 May;94(5):647-652.)
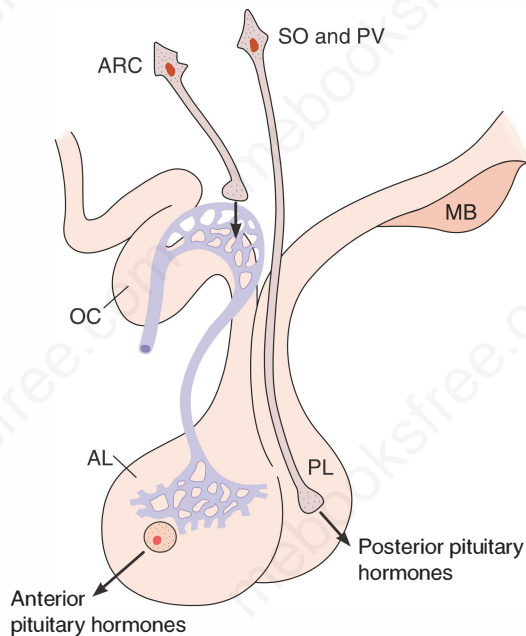
FIGURE 4–5   Secretion of hypothalamic hormones. The hormones of the posterior lobe (PL) are released into the general circulation from the endings of supraoptic and paraventricular neurons, whereas hypophysiotropic hormones are secreted into the portal hypophysial circulation from the endings of arcuate and other hypothalamic neurons (AL, anterior lobe; ARC, arcuate and other nuclei; MB, mamillary bodies; OC, optic chiasm; PV, paraventricular nucleus; SO, supraoptic nucleus).

cells emerge in a specific order and from distinct lineages. Abnormalities of pituitary and lineage–specific transcription factors have been associated with the development of hypopituitarism. Although traditionally the pituitary has been conceptualized as a gland with distinct and highly specialized cells that respond to specific hypothalamic and peripheral hormones, it has become clear that local (ie, paracrine) factors also play a role in normal pituitary physiology.

A. Somatotrophs   The GH-secreting cells are acidophilic and usually located in the lateral portions of the anterior lobe. Granule size by electron microscopy is 150 to 600 nm in diameter. These cells account for about 50% of the adenohypophysial cells.

B. Lactotrophs   The PRL-secreting cell is a second but distinct acidophil-staining cell randomly distributed in the anterior pituitary. These cells account for 10% to 25% of anterior pituitary cells. Granule size averages approximately 550 nm on electron microscopy. There are two types of lactotrophs: sparsely granulated and densely granulated. These cells proliferate during pregnancy as a result of elevated estrogen levels and account for the twofold increase in gland size.

C. Thyrotrophs   These TSH-secreting cells, because of their glycoprotein product, are basophilic and also show a positive reaction with periodic acid-Schiff stain. Thyrotrophs are the least common pituitary cell type, making up less than 10% of adenohypophysial cells. The thyrotroph granules are small (50-100 nm), and these cells are usually located in the anteromedial
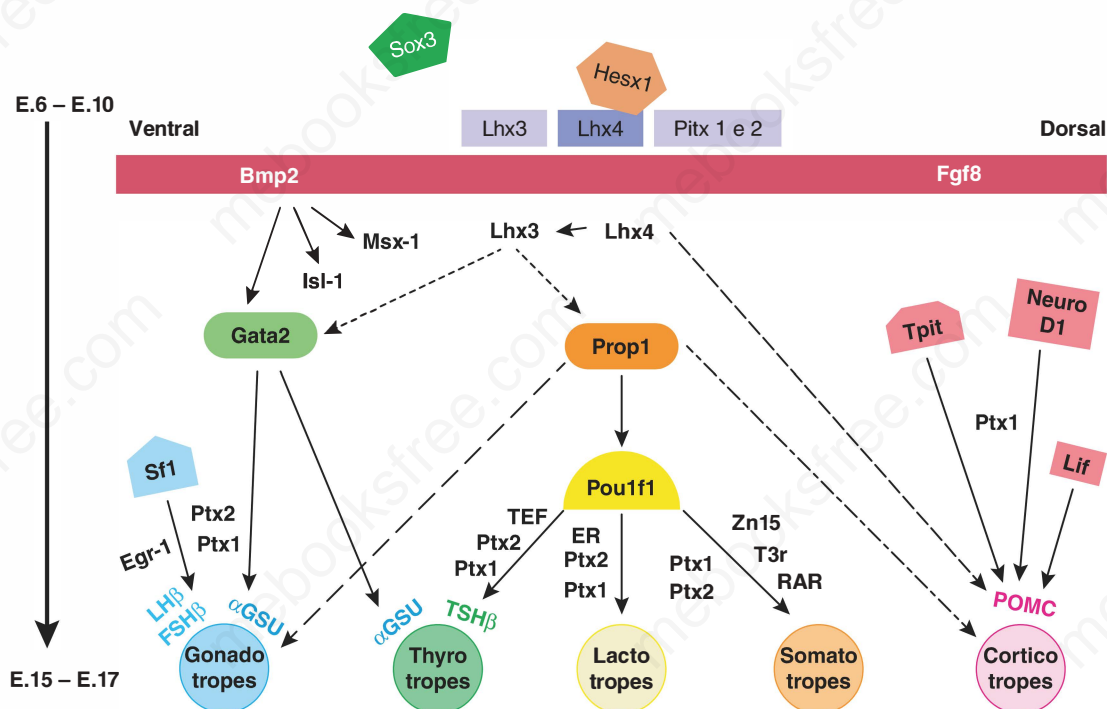
FIGURE 4–6 Temporal and spatial activation of pituitary transcription factors. In response to the BMP2-FGF8 ventral-dorsal gradient, pituitary cell lineages are determined by the activation or repression of each transcription factor. Solid arrows indicate the activation of expression, dotted arrows indicate an unknown role in the activation of expression, dashed arrows indicate an undefined role and dash-dot arrows indicate an action of an important factor in the maintenance of long-term cell function (Bmp2, bone morphogenic protein 2; Egr1, early growth response 1; ER, estrogen receptor; Fgf8, fibroblast growth factor 8; Gata2, GATA-binding protein 2; Hesx 1, HESX homeobox 1; Isl1, ISL LIM homeobox 1; Lhx3, LIM homeobox 3; Lhx4, LIM homeobox 4; Lif, leukemia inhibitory factor; Msx1, msh homeobox 1; NeuroD1, neurogenic differentiation 1; PIT1, POU class 1 homeobox 1; Pitx1, paired-like homeodomain 1; Pitx2, paired-like homeodomain 2; POMC, pro-opiomelanocortin; Prop1, prophet of Pit-1; RAR, retinoic acid receptor; SF1, steroidogenic factor 1; T3r, thyroid hormone nuclear receptor; TEF, thyrotrope embryonic factor; Tpit, T-box 19; Zn15, zinc finger protein Zn15). (Reproduced with permission from de Moraes et al. Pituitary development: a complex, temporal regulated process dependent on specific transcriptional factors. *J Endocrinol*. 2012 Nov;215(2):239-245.)

and anterolateral portions of the gland. During states of primary thyroid failure, the cells demonstrate marked hypertrophy, increasing overall gland size.

D. Corticotrophs   ACTH and its related peptides (see later) are secreted by basophilic cells that are embryologically of intermediate lobe origin and usually located in the anteromedial portion of the gland. Corticotrophs represent 15% to 20% of adenohypophysial cells. Electron microscopy shows that these secretory granules are about 360 nm in diameter. In states of glucocorticoid excess, corticotrophs undergo degranulation and a microtubular hyalinization known as Crooke hyaline degeneration (Figure 4–7).

E. Gonadotrophs   LH and FSH originate from basophil-staining cells, whose secretory granules are about 200 nm in diameter. These cells constitute 10% to 15% of anterior pituitary cells, and they are located throughout the entire anterior lobe. They become hypertrophied and cause the gland to enlarge during states of primary gonadal failure such as menopause, Klinefelter syndrome, and Turner syndrome.
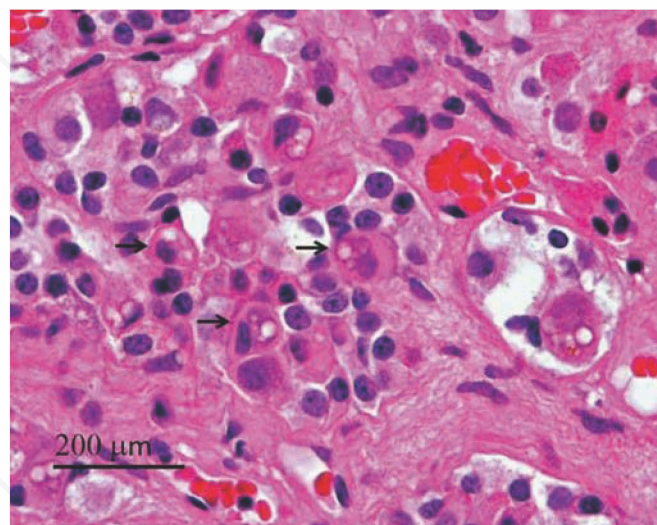
FIGURE 4–7 Crooke hyaline change in non-neoplastic corticotrophs of a patient with Cushing syndrome (H&E stain 40× magnification). The arrows denote perinuclear aggregates of cytokeratin (bright hyaline material surrounding the nuclei and vacuoles). (Used with permission from Beatriz Lopes, MD, PhD.)

TABLE 4–2 Major adenohypophysial hormones and their cellular sources.

Cellular Source and Histologic Staining	Main Hormone Products	Structure of Hormone	Main Functions
Somatotroph (acidophil)	GH; also known as STH or somatotropin	191 amino acids, 22-kDa protein, mainly nonglycosylated	Stimulates the production of IGF-1 (the mediator of the indirect actions of GH); also exerts direct actions on growth and metabolism
Lactotroph or mammotroph (acidophil)	PRL	198 amino acids, 23-kDa protein, mainly nonglycosylated (Note: most of the decidually produced PRL is glycosylated)	Stimulation of milk production (protein and lactose synthesis, water excretion, and sodium retention); inhibits gonadotropin; immunomodulator
Corticotroph (small cells with basophil granules with strong PAS positivity, indicating the presence of glycoproteins)	Derivatives of POMC, mainly ACTH and β-LPH	POMC: glycosylated polypeptide of 134 amino acid residues ACTH: simple peptide of 39 amino acid residues, 4.5 kDa β-LPH: simple peptide of 91 amino acid residues, 11.2 kDa	ACTH: stimulation of glucocorticoids and sex steroids in the zona fasciculata and zona reticularis of the adrenal cortex, inducing hyperplasia and hypertrophy of the adrenal cortex β-LPH: weak lipolytic and opioid actions
Thyrotroph (large cells with basophil granules with PAS positivity)	TSH	Glycoprotein hormone consisting of a shared α (89 amino acid) and a TSH-specific β (112 amino acid) subunit Total size: 28 kDa	Stimulation of all aspects of thyroid gland function: hormone synthesis, secretion, hyperplasia, hypertrophy, and vascularization
Gonadotroph (small cells with basophil granules with periodic acid-Schiff positivity)	LH: named after its effect in females; is identical to the ICSH (interstitial cell stimulating hormone) originally described in males	Glycoprotein hormone consisting of a shared α and an LH-specific β (115 amino acid) subunit Total size: 29 kDa	Females: stimulates steroid hormone synthesis in theca interna cells, lutein cells, and hilar cells; promotes luteinization and maintains corpus luteum Males: stimulates steroid hormone production in Leydig cells
	FSH	Glycoprotein hormone consisting of a shared α and an FSH-specific β (115 amino acid) subunit Total size: 29 kDa	Females: targets the granulosa cells to promote follicular development; stimulates aromatase expression and inhibin secretion Males: targets the Sertoli cells to promote spermatogenesis and to stimulate inhibin secretion

Modified with permission from Kacsoh B. *Endocrine Physiology*. New York: McGraw-Hill Education; 2000.

F. Other cell types   Some cells, usually chromophobes, contain secretory granules but do not exhibit immunocytochemical staining for the major known anterior pituitary hormones. These cells have been called null cells; they may give rise to (apparently) nonfunctioning adenomas. Some may represent undifferentiated primitive secretory cells, and others (eg, glia-like or folliculostellate cells) may produce one or more of the many paracrine factors that have been described in the pituitary. Mammosomatotrophs contain both GH and PRL; these bihormonal cells are most often seen in pituitary tumors. Human chorionic gonadotropin (hCG) is also secreted by the anterior pituitary gland, but its cell of origin and physiologic significance are uncertain. The six known major anterior pituitary hormones are listed in Table 4–2.

HYPOTHALAMIC HORMONES

The hypothalamic hormones can be divided into those secreted into hypophysial portal blood vessels and those secreted by the neurohypophysis directly into the general circulation. The hypothalamic nuclei, their neurohormones, and their main functions

are shown in Table 4–3. The structures of the eight major hypothalamic hormones are shown in Table 4–4.

Hypophysiotropic Hormones

The hypophysiotropic hormones that regulate the secretion of anterior pituitary hormones include growth hormone–releasing hormone (GHRH), somatostatin (growth hormone–inhibiting hormone [GHIH]), dopamine, thyrotropin-releasing hormone (TRH), corticotropin-releasing hormone (CRH), and gonadotropin-releasing hormone (GnRH). The location of the cell bodies of the hypophysiotropic hormone–secreting neurons is depicted in Figure 4–8. Most of the anterior pituitary hormones are controlled by stimulatory hormones, but GH and especially PRL are also regulated by inhibitory hormones. Some hypophysiotropic hormones are multifunctional. The hormones of the hypothalamus are secreted episodically and not continuously, and in some cases there is an underlying circadian rhythm.

A. GHRH   GHRH stimulates GH secretion by, and is trophic for, somatotrophs. GHRH-secreting neurons are located in the arcuate

TABLE 4–3 The hypothalamic nuclei and their main functions.

Nucleus	Location	Major Neurohormones or Functions
Supraoptic	Anterolateral, above the optic tract	ADH: osmoregulation, regulation of ECF volume Oxytocin: regulation of uterine contractions and milk ejection
Paraventricular	Dorsal anterior periventricular	Magnocellular paraventricular nucleus (PVN): ADH, oxytocin: same functions as above Parvocellular PVN TRH: regulation of thyroid function CRH: regulation of adrenocortical function, regulation of the sympathetic nervous system and adrenal medulla, regulation of appetite ADH: coexpressed with CRH, regulation of adrenocortical function VIP: prolactin-releasing factor (?)
Suprachiasmatic	Above the optic chiasm, anteroventral periventricular zone	Regulator of circadian rhythms and pineal function (zeitgeber [pacemaker]): VIP, ADH neurons project mainly to the PVN
Arcuate	Medial basal hypothalamus close to the third ventricle	GHRH: stimulation of growth hormone GnRH: regulation of pituitary gonadotropins (FSH and LH) Dopamine: functions as PIH Somatostatin: inhibition of GHRH release Regulation of appetite (neuropeptide Y, agouti-related transcript, α-MSH, cocaine- and amphetamine-related transcript)
Periventricular	Anteroventral	Somatostatin: inhibition of growth hormone secretion by direct pituitary action: most abundant SRIF location
Ventromedial	Ventromedial	GHRH (as above) Somatostatin: inhibition of GHRH release Functions as a satiety center
Dorsomedial	Dorsomedial	Focal point of information processing: receives input from ventromedial nucleus (VMN) and lateral hypothalamus and projects to the PVN
Lateral hypothalamus	Lateral hypothalamus	Functions as a hunger center (melanin-concentrating hormone, anorexins)
Preoptic area	Preoptic area	Main regulator of ovulation in rodents; only a few GnRH neurons in primates
Anterior hypothalamus	Anterior hypothalamus	Thermoregulation: cooling center Anteroventral third ventricular region: regulation of thirst
Posterior hypothalamus	Posterior hypothalamus	Thermoregulation: heating center

Modified with permission from Kacsoh B. *Endocrine Physiology*. New York: McGraw-Hill Education; 2000.

nuclei (see Figure 4–2), and axons terminate in the external layer of the median eminence. The major isoform of GHRH is 44 amino acids in length. It was isolated from a pancreatic tumor in a patient with clinical manifestations of GH excess (acromegaly) associated with somatotroph hyperplasia (see discussion later in the chapter). GHRH is synthesized from a larger precursor of 108 amino acids. Other secretory products derived from this precursor have also been found. Full biologic activity of these releasing factors appears to reside in the 1 to 29 amino acid sequence of the amino terminal portion of the molecule. Human GHRH is a member of a homologous family of peptides that includes secretin, glucagon, vasoactive intestinal peptide (VIP), and others. The half-life of GHRH is approximately 3 to 7 minutes.
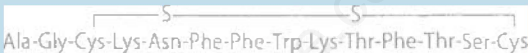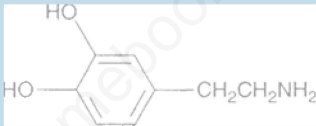
B. Somatostatin   Somatostatin inhibits the secretion of GH and TSH. Somatostatin-secreting cells are located in the periventricular region immediately above the optic chiasm (see Figure 4–2) with nerve endings found diffusely in the external layer of the median eminence.

Somatostatin, a tetradecapeptide, has been found not only in the hypothalamus but also in the D cells of the pancreatic

islets, the gastrointestinal mucosa, and the C cells (parafollicular cells) of the thyroid. The somatostatin precursor has 116 amino acids. Processing of the carboxyl terminal region of preprosomatostatin results in the generation of the tetradecapeptide somatostatin 14 and an amino terminal extended form containing 28 amino acid residues (somatostatin 28). Somatostatin 14 is the major species in the hypothalamus, whereas somatostatin 28 is found in the gut. In addition to its profound inhibitory effect on GH secretion, somatostatin also has important inhibitory influences on many other hormones, including insulin, glucagon, gastrin, secretin, and VIP. This inhibitory hypothalamic peptide plays a role in the physiologic secretion of TSH by augmenting the direct inhibitory effect of thyroid hormone on the thyrotrophs; administration of antisomatostatin antibodies results in a rise in circulating TSH level. Somatostatin has a half-life of 2 to 3 minutes.

C. Dopamine   Dopamine, the primary PRL-inhibitory hormone, is found in the portal circulation and binds to dopamine receptors on lactotrophs. It has a short half-life, on the order of 1 to 2 minutes. The hypothalamic control of PRL secretion, unlike

TABLE 4–4 Hypothalamic hormones.

Hormone	Structure
Posterior pituitary hormones	
Arginine vasopressin	Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly-NH$_2$ (disulfide bridge Cys–Cys)
Oxytocin	Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH$_2$ (disulfide bridge Cys–Cys)
Hypophysiotropic hormones	
Thyrotropin-releasing hormone (TRH)	(pyro)Glu-His-Pro-NH$_2$
Gonadotropin-releasing hormone (GnRH)	(pyro)Glu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH$_2$
Somatostatin[a] (growth hormone-inhibiting hormone [GHIH])	Ala-Gly-Cys-Lys-Asn-Phe-Phe-Trp-Lys-Thr-Phe-Thr-Ser-Cys (disulfide bridge Cys–Cys)
Growth hormone–releasing hormone (GHRH)	Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Met-Ser-Arg-Gln-Gln-Gly-Glu-Ser-Asn-Gln-Glu-Arg-Gly-Ala-Arg-Ala-Arg-Leu-NH$_2$
Prolactin-inhibiting hormone (PIH, dopamine)	
Corticotropin-releasing hormone (CRH)	Ser-Gln-Glu-Pro-Pro-Ile-Ser-Leu-Asp-Leu-Thr-Phe-His-Leu-Leu-Arg-Glu-Val-Leu-Glu-Met-Thr-Lys-Ala-Asp-Gln-Leu-Ala-Gln-Gln-Ala-His-Ser-Asn-Arg-Lys-Leu-Leu-Asp-Ile-Ala-NH$_2$

[a]In addition to the tetradecapeptide shown here (somatostatin 14), an amino terminal–extended molecule (somatostatin 28), and a 12-amino-acid form (somatostatin 28 [1-12]) are found in most tissues.

that of the other pituitary hormones, is predominantly inhibitory. Thus, disruption of the hypothalamic-pituitary connection by stalk section, hypothalamic lesions, or pituitary autotransplantation increases PRL secretion. Dopamine-secreting neurons (tuberoinfundibular dopaminergic system) are located in the arcuate nuclei, and their axons terminate in the external layer of the median eminence, primarily in the same area as the GnRH endings (laterally) and to a lesser extent medially (see Figure 4–2). The neurotransmitter gamma-aminobutyric acid (GABA) and cholinergic pathways also appear to inhibit PRL release.

D. Prolactin-releasing factors   The best-studied factor with PRL-releasing activity is TRH (see discussion later), but there is little evidence for a physiologic role. PRL increase associated with sleep, during stress, and after nipple stimulation or suckling is not accompanied by an increase in TRH or TSH. Another hypothalamic peptide, VIP, stimulates PRL release in humans. Serotonergic pathways may also stimulate PRL secretion, as demonstrated by the increased PRL secretion after the administration of serotonin precursors and by the reduction of secretion following treatment with serotonin antagonists.

E. Thyrotropin-releasing hormone   TRH, a tripeptide, is the major hypothalamic factor regulating TSH secretion. Human TRH is synthesized from a large precursor of 242 amino acids that contains six copies of TRH. TRH-secreting neurons are located in the medial portions of the paraventricular nuclei (see Figure 4–2), and their axons terminate in the medial portion of the external layer of the median eminence. The half-life of TRH is approximately 6 minutes.

F. Corticotropin-releasing hormone   CRH, a 41-amino-acid peptide, stimulates the secretion of ACTH and other products of its precursor molecule, POMC. CRH is synthesized from a precursor of 196 amino acids. The half-life of CRH follows a biphasic pattern in plasma lasting approximately 6 to 10 minutes and 40 to 50 minutes. Both antidiuretic hormone (ADH) and angiotensin II potentiate CRH-mediated secretion of ACTH. In contrast, oxytocin inhibits CRH-mediated ACTH secretion. Opioids appear to suppress release of CRH. CRH-secreting neurons are found in the anterior portion of the paraventricular nuclei just lateral to the TRH-secreting neurons; their nerve endings are found in all parts of the external layer of the median eminence. CRH is also secreted from human placenta. The level of this hormone increases significantly during late pregnancy and delivery. In addition, a specific CRH-binding protein (CRHBP) has been described in both serum and in intracellular locations within a variety of cells. It is likely that CRHBPs modulate the actions and plasma half-life of CRH. Since the 1990s, three proteins homologous to CRH, termed urocortins, and two different receptors, have been identified. In addition to the role of CRH in the physiologic response to stress, this family of peptides appears to play a significant role in energy balance.

G. Gonadotropin-releasing hormone   The secretion of LH and FSH is controlled by a single stimulatory hypothalamic
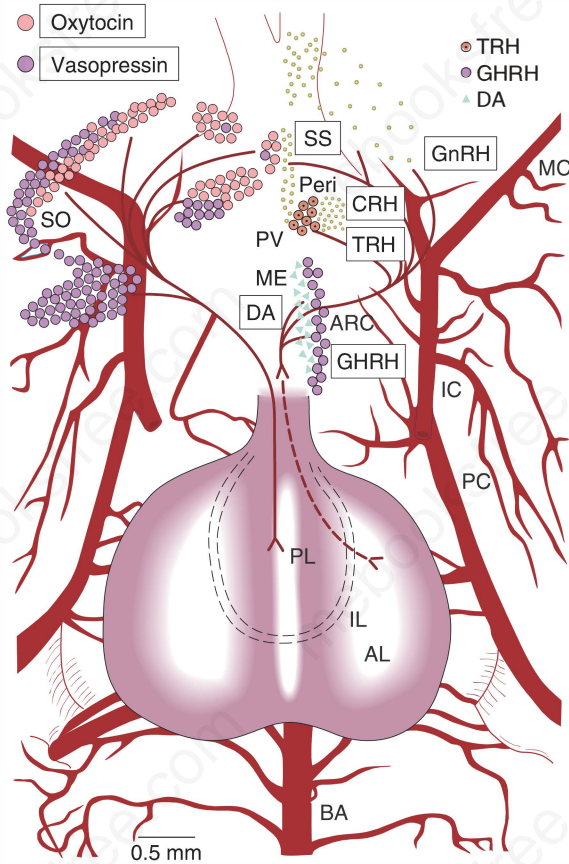
FIGURE 4–8 Location of cell bodies of hypophysiotropic hormone-secreting neurons projected on a ventral view of the hypothalamus and pituitary of the rat (AL, anterior lobe; ARC, arcuate nucleus; BA, basilar artery; IC, internal carotid; IL, intermediate lobe; MC, middle cerebral; ME, median eminence; PC, posterior cerebral; Peri, periventricular nucleus; PL, posterior lobe; PVL and PVM, lateral and medial portions of the paraventricular nucleus; SO, supraoptic nucleus). The names of the hormones are enclosed in the boxes (SS, somatostatin; DA, dopamine). (Used with permission from LW Swanson and ET Cunningham Jr.)

Neuroendocrinology: The Hypothalamus as Part of a Larger System

The hypothalamus is involved in many nonendocrine functions such as regulation of body temperature, thirst, and food intake and is connected with many other parts of the nervous system. The brain itself is influenced by both direct and indirect hormonal effects. Steroid and thyroid hormones cross the blood–brain barrier and produce specific receptor-mediated actions (see Chapters 1, 7, and 9). Peptides in the general circulation, which do not cross the blood–brain barrier, elicit their effects indirectly (eg, insulin-mediated changes in blood–glucose concentration). In addition, communication between the general circulation and the brain may take place via the circumventricular organs, which are located outside the blood–brain barrier (see later). Moreover, hypothalamic hormones in extrahypothalamic brain function as neurotransmitters or neurohormones. They are also found in other tissues where they function as hormones (endocrine, paracrine, or autocrine). For example, somatostatin-containing neurons are widely distributed in the nervous system. They are also found in the pancreatic islets (D cells), the gastrointestinal mucosa, and the C cells of the thyroid gland (parafollicular cells). Somatostatin is not only secreted into the general circulation as well as locally—it is also secreted into the lumen of the gut, where it may affect gut secretion. A hormone with this activity has been called a **lumone**. Hormones common to the brain, pituitary, and gastrointestinal tract include not only TRH and somatostatin but also VIP and peptides derived from POMC.

Hypothalamic function is regulated both by hormone-mediated signals (eg, negative feedback) and by neural inputs from a wide variety of sources. These nerve signals are mediated by neurotransmitters including acetylcholine, dopamine, norepinephrine, epinephrine, serotonin, GABA, and opioids. The hypothalamus can be considered a final common pathway by which signals from multiple systems reach the anterior pituitary. For example, cytokines that play a role in the response to infection, such as the interleukins, are also involved in regulation of the hypothalamic-pituitary-adrenal axis. This system of immunoneuroendocrine interactions is important in the organism's response to a variety of stresses.

The hypothalamus also sends signals to other parts of the nervous system. For example, while the major nerve tracts of the magnocellular neurons containing vasopressin and oxytocin terminate in the posterior pituitary, nerve fibers from the paraventricular and supraoptic nuclei project toward many other parts of the nervous system. In the brain stem, vasopressinergic neurons are involved in the autonomic regulation of blood pressure. Similar neurons project to the gray matter and are implicated in higher cortical functions. Fibers terminating in the median eminence permit release of ADH into the hypophysial-portal system; delivery of ADH in high concentrations to the anterior pituitary may facilitate its involvement in the regulation of ACTH secretion. Magnocellular neurons also project to the choroid plexus, where they may release ADH into the cerebrospinal fluid. In addition to magnocellular neurons, the

hormone, GnRH. This is achieved through differences in the size and frequency of GnRH release as well as feedback from estrogens and androgens; low-frequency pulses favor FSH release while high-frequency pulses result in LH release. GnRH is a linear decapeptide that stimulates only LH and FSH; it has no effect on other pituitary hormones except in some patients with acromegaly and Cushing disease (see discussion later). The precursor of GnRH—proGnRH—contains 92 amino acids. ProGnRH also contains the sequence of a 56-amino-acid polypeptide called GnRH-associated peptide. This secretory product exhibits PRL-inhibiting activity, but its physiologic role is unknown. GnRH-secreting neurons are located primarily in the preoptic area of the anterior hypothalamus, and their nerve terminals are found in the lateral portions of the external layer of the median eminence adjacent to the pituitary stalk (see Figure 4–2). GnRH has a half-life of 2 to 4 minutes.

paraventricular nuclei contain cells with smaller cell bodies—**parvicellular neurons**. Such neurons are also found in other regions of the nervous system and may contain other peptides such as CRH and TRH.

The Hypothalamus and the Control of Appetite

With the growing appreciation of adipose tissue as an endocrine organ, as well as the increasing problem of obesity and its associated health risks, understanding how energy balance and appetite are regulated has become a major topic of study. In 1901, Frohlich observed that some tumors affecting the pituitary and hypothalamus were associated with excess subcutaneous fat and hypogonadism. Subsequent lesioning experiments by Hetherington and Ranson in the 1940s established the hypothalamus as a fundamental site in the regulation of appetite. These experiments introduced the classic *dual center* model of food intake where the ventromedial hypothalamic nucleus functions as a *satiety center* and the lateral hypothalamic area serves as a *feeding center*. Subsequent studies have led to refinements of this model.

Growing evidence points to the pivotal role of the arcuate nucleus in the integration of feeding signals and energy reserves. It has special access to circulating hormones via the underlying median eminence, an area rich with fenestrated capillaries, that is, it is not protected by the blood–brain barrier (see Figure 4–2). Two populations of neurons involved in the regulation of feeding are located within the arcuate nucleus: one that inhibits food intake via the expression of neuropeptides POMC and cocaine- and amphetamine-regulated transcript, and one that stimulates food intake via expression of neuropeptide Y and agouti-related peptide. The arcuate nucleus projects to second-order neuronal populations including the paraventricular nucleus, dorsomedial nucleus, ventromedial nucleus, and lateral hypothalamic area, which then activate downstream pathways controlling appetite and energy expenditure.

Circulating markers of adiposity (leptin, adiponectin, insulin), and hormones from the gastrointestinal tract (ghrelin, peptide YY, glucagon-like peptide 1, cholecystokinin, oxyntomodulin, pancreatic polypeptide) converge on the hypothalamus and brain stem to signal adequacy of short- and long-term energy stores. Alterations in the levels or tissue sensitivity of these hormones may underlie disorders of weight regulation such as obesity, and could prove useful as therapeutic targets (see also Chapter 20).

The Pineal Gland and the Circumventricular Organs

The circumventricular organs are secretory midline brain structures that arise from the ependymal cell lining of the ventricular system (Figure 4–9). These organs are located adjacent to the third ventricle—subfornical organ, subcommissural organ, organum vasculosum of the lamina terminalis, pineal, and part of the median eminence—and at the roof of the fourth ventricle—area postrema (see Figure 4–9). The tissues of these organs have relatively large interstitial spaces and have fenestrated capillaries
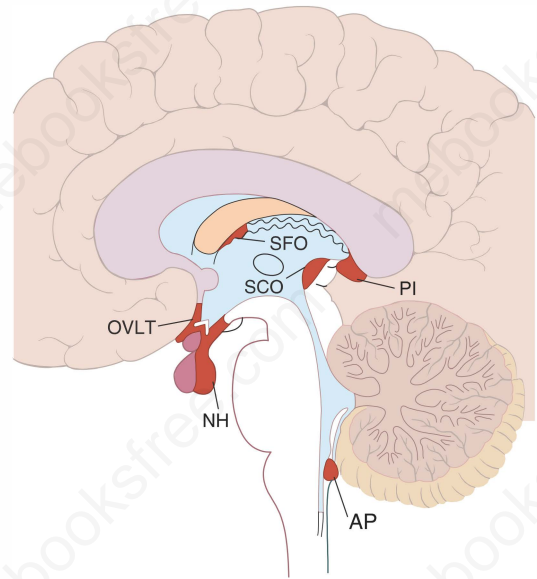
FIGURE 4–9 Circumventricular organs. The neurohypophysis (NH) and adjacent median eminence, the organum vasculosum of the lamina terminalis (OVLT), the subfornical organ (SFO), and the area postrema (AP) are shown projected on a sagittal section of the human brain. The pineal (PI) and the subcommissural organ (SCO) are also shown but probably do not function as circumventricular organs. (Reproduced with permission from Ganong WF. *Review of Medical Physiology*. 15th ed. New York: McGraw-Hill Education; 1993.)

which, being highly permeable, permit diffusion of large molecules from the general circulation; elsewhere in the brain tight capillary endothelial junctions prevent such diffusion—the blood–brain barrier. For example, angiotensin II (see Chapter 10) is involved in the regulation of water intake, blood pressure, and secretion of vasopressin. In addition to its peripheral effects, circulating angiotensin II acts on the subfornical organ resulting in an increase in water intake.

The pineal gland, considered by the 17th century French philosopher René Descartes to be the seat of the soul, is located at the roof of the posterior portion of the third ventricle. The pineal gland in humans and other mammals has no direct neural connections with the brain except for sympathetic innervation via the superior cervical ganglion. The pineal gland secretes melatonin, an indole synthesized from serotonin by 5-methoxylation and *N*-acetylation (Figure 4–10). The pineal releases melatonin into the general circulation and into the cerebrospinal fluid. Melatonin secretion is regulated by the sympathetic nervous system and is increased in response to hypoglycemia and darkness. The pineal also contains other bioactive peptides and amines including TRH, somatostatin, GnRH, and norepinephrine. The physiologic roles of the pineal remain to be elucidated, but they appear to involve regulation of gonadal function and development and chronobiologic rhythms.

The pineal gland may be the site of origin of pineal cell tumors (pinealomas) or germ cell tumors (germinomas). Neurologic signs and symptoms are the predominant clinical manifestations; examples include increased intracranial pressure, visual
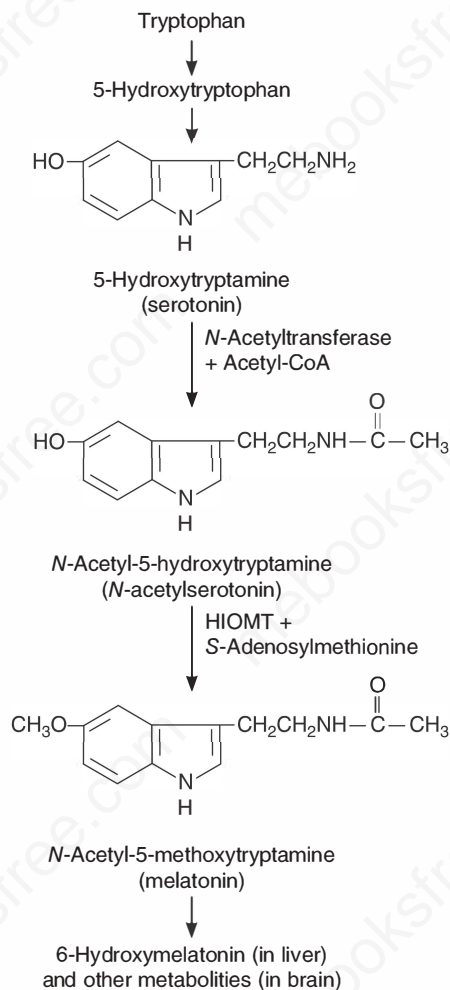
FIGURE 4–10 Formation and metabolism of melatonin (HIOMT, hydroxyindole-O-methyltransferase.) (Reproduced with permission from Ganong WF. *Review of Medical Physiology*. 15th ed. New York: McGraw-Hill Education; 1993.)

abnormalities, ataxia, and Parinaud syndrome—upward gaze palsy, absent pupillary light reflex, paralysis of convergence, and wide-based gait. Endocrine manifestations result primarily from deficiency of hypothalamic hormones (diabetes insipidus, hypopituitarism, or disorders of gonadal development). Treatment involves surgical removal or decompression, radiation therapy, and hormone replacement (discussed later).

ANTERIOR PITUITARY HORMONES

The six major anterior pituitary hormones—ACTH, GH, PRL, TSH, LH, and FSH—may be classified into three groups: ACTH-related peptides (ACTH itself, β-lipotropin [β-LPH], melanocyte-stimulating hormone [MSH], and endorphins); the somatomammotropin proteins (GH and PRL); and the glycoproteins (LH, FSH, and TSH). The chemical features of these hormones are set forth in Table 4–2.

ADRENOCORTICOTROPIC HORMONE AND RELATED PEPTIDES

Biosynthesis

ACTH is a 39-amino-acid peptide hormone (MW 4500) processed from a large precursor molecule, POMC (MW 28,500). Within the corticotroph, a single mRNA directs the synthesis and processing of POMC into smaller, biologically active fragments (Figure 4–11), which include β-LPH, α-MSH, β-MSH, β-endorphin, and the amino terminal fragment of POMC. Most of these peptides are glycosylated, which accounts for differences in the reporting of molecular weights. These carbohydrate moieties are responsible for the basophilic staining of corticotrophs.
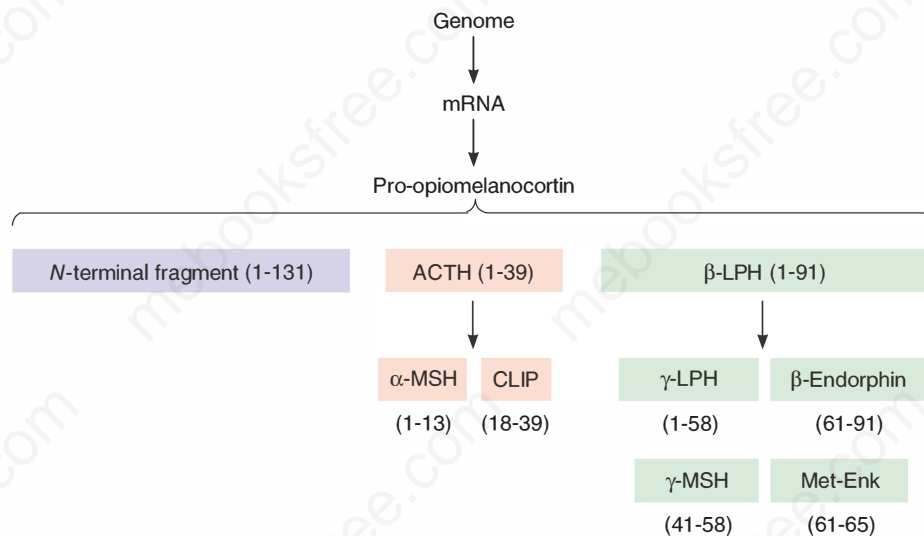
FIGURE 4–11 The processing of pro-opiomelanocortin (MW 28,500) into its biologically active peptide hormones. Abbreviations are expanded in the text.

Two of these fragments are contained within the structure of ACTH: α-MSH is identical to ACTH 1 to 13, and corticotropin-like intermediate lobe peptide (CLIP) represents ACTH 18 to 39 (see Figure 4–11). Although these fragments are found in species with developed intermediate lobes (eg, the rat), they are not secreted as separate hormones in humans. β-LPH, a fragment with 91 amino acids (1-91), is secreted by the corticotroph in equimolar quantities with ACTH. Within the β-LPH molecule exists the amino acid sequence for β-MSH (41-58), γ-LPH (1-58), and β-endorphin (61-91).

Function

ACTH stimulates the secretion of glucocorticoids, mineralocorticoids, and androgens—all steroids from the adrenal cortex (see Chapters 9 and 10). The amino terminal end (residues 1-18) is responsible for this biologic activity. ACTH binds to receptors on the adrenal cortex and induces steroidogenesis through a cAMP-dependent mechanism.

The hyperpigmentation observed in states of ACTH hypersecretion (eg, Addison disease, Nelson syndrome) appears to be primarily due to ACTH binding to the MSH receptor, because α-MSH and β-MSH do not exist as separate hormones in humans.

The physiologic function of β-LPH and its family of peptide hormones, including β-endorphin, is not completely understood. However, both β-LPH and β-endorphin have the same secretory dynamics as ACTH.

Measurement

The development of immunoradiometric and immunochemiluminescent assays (IRMAs and ICMAs, respectively) has provided a sensitive and practical clinical ACTH assay for the evaluation of pituitary-adrenal disorders. The basal morning concentration ranges from 9 to 52 pg/mL (2-11 pmol/L). Its short plasma half-life (7-12 minutes) and episodic secretion cause wide and rapid fluctuations both in its plasma concentration and in that of cortisol.

Although β-LPH has a longer half-life than ACTH and is more stable in plasma, its measurement has not been extensively utilized. Data suggest that the normal concentration of β-LPH is 10 to 40 pg/mL (1-4 pmol/L).

Secretion

The physiologic secretion of ACTH is mediated through neural influences by means of a complex of hormones, the most important of which is CRH (Figure 4–12).

CRH stimulates ACTH in a pulsatile manner: diurnal rhythmicity causes a peak before awakening and a decline as the day progresses. The diurnal rhythm is a reflection of neural control and provokes concordant diurnal secretion of cortisol from the adrenal cortex (Figure 4–13). This episodic release of ACTH is independent of circulating cortisol levels (ie, the magnitude of an ACTH impulse is not related to preceding plasma cortisol levels). An example is the persistence of diurnal rhythm in patients with primary adrenal failure (Addison disease). ACTH secretion also increases in response to feeding in both humans and animals.
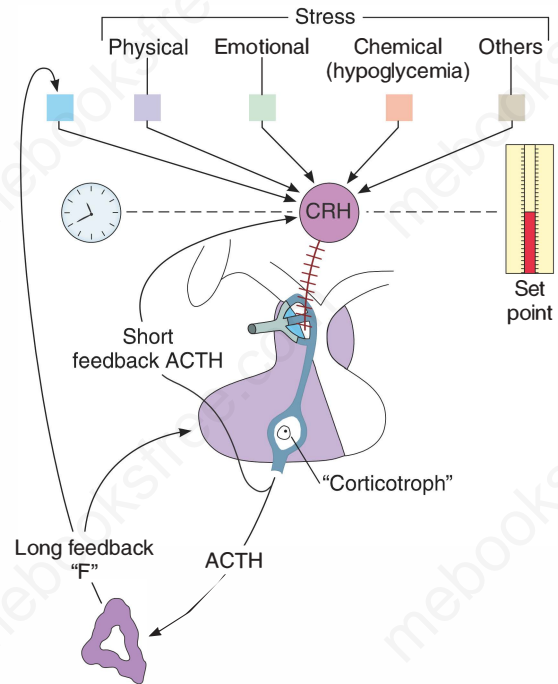
FIGURE 4–12 The hypothalamic-pituitary-adrenal axis, illustrating negative feedback by cortisol (F) at the hypothalamic and pituitary levels. A short negative feedback loop of ACTH on the secretion of corticotropin-releasing hormone (CRH) also exists. (Reproduced with permission from Gwinup G, Johnson B. Clinical testing of the hypothalamic-pituitary-adrenocortical system in states of hypo- and hypercortisolism. *Metabolism*. 1975 Jun;24(6):777-791.)

Many stresses stimulate ACTH, often superseding the normal diurnal rhythmicity. Physical, emotional, and chemical stresses such as pain, trauma, hypoxia, acute hypoglycemia, cold exposure, surgery, depression, and interleukin-1 and vasopressin administration have all been shown to stimulate ACTH and cortisol secretion. The increase in ACTH levels during stress is mediated by vasopressin as well as CRH. Although physiologic cortisol levels do not blunt the ACTH response to stress, exogenous corticosteroids in high doses suppress it.

Negative feedback of cortisol and synthetic glucocorticoids on ACTH secretion occurs at both the hypothalamic and pituitary levels via two mechanisms: fast feedback is sensitive to the rate of change in cortisol levels, whereas slow feedback is sensitive to the absolute cortisol level. The first mechanism is probably nonnuclear; that is, this phenomenon occurs too rapidly to be explained by the influence of corticosteroids on nuclear transcription of the specific mRNA responsible for ACTH. Recent studies suggest fast feedback is mediated by a novel membrane-associated glucocorticoid receptor that stimulates a rapid synthesis and retrograde release of endocannabinoids, thereby suppressing synaptic excitation. Slow feedback, occurring later, may be explained by a nuclear-mediated mechanism and a subsequent decrease in synthesis of ACTH. This latter form of negative feedback is the type probed by the clinical dexamethasone suppression test. In addition to the negative feedback of corticoids, ACTH also inhibits its own secretion (short-loop feedback).
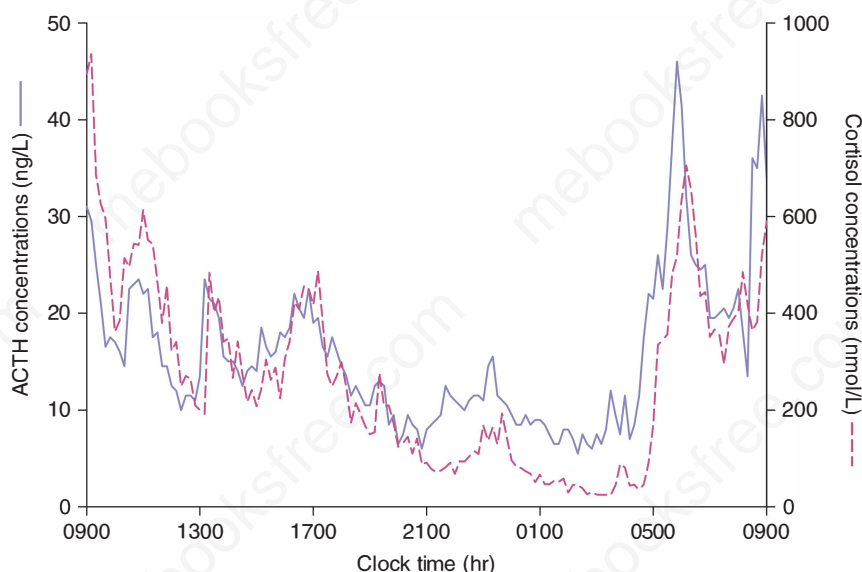
FIGURE 4–13 The episodic, pulsatile pattern of ACTH secretion and its concordance with cortisol secretion in a healthy human subject during a 24-hour period. (Used with permission from Johannes D. Veldhuis, MD.)

GROWTH HORMONE

Biosynthesis

GH or somatotropin is a 191-amino-acid polypeptide hormone (MW 21,500) synthesized and secreted by the somatotrophs of the anterior pituitary. Its larger precursor peptide, pre-GH (MW 28,000), is also secreted but has no physiologic significance.

Function

The primary function of GH is promotion of linear growth. Its basic metabolic effects serve to achieve this result, but most of the growth-promoting effects are mediated by insulin-like growth factor 1 (IGF-1; previously known as somatomedin C). The metabolic and biologic effects of GH and IGF-1 are shown in Tables 4–5 and 4–6 (see also Chapter 6).

GH, via IGF-1, increases protein synthesis by enhancing amino acid uptake and directly accelerating the transcription and translation of mRNA. In addition, GH tends to decrease protein catabolism by mobilizing fat as a more efficient fuel source: it directly causes the release of fatty acids from adipose tissue and enhances their conversion to acetyl-CoA, from which energy is derived. This protein-sparing effect is an important mechanism by which GH promotes growth and development.

GH also affects carbohydrate metabolism. In excess, it decreases carbohydrate utilization and impairs glucose uptake into cells. This GH-induced insulin resistance appears to be due to a postreceptor impairment in insulin action. These events result in glucose intolerance and secondary hyperinsulinism.

Measurement

GH has a plasma half-life of 10 to 20 minutes. The healthy adult secretes approximately 400 µg/d (18.6 nmol/d); in contrast, young adolescents secrete about 700 µg/d (32.5 nmol/d). The early morning GH concentration in fasting unstressed adults is less than 2 ng/mL (90 pmol/L). There are no significant sex differences.

About half of circulating GH is bound to specific GH-binding proteins (GHBPs) that function to reduce oscillations in GH levels (due to its pulsatile secretion) and prolong plasma GH half-life. GHBPs include a high-affinity GHBP (corresponding to the extracellular portion of the GH receptor formed through

TABLE 4–5 Metabolic effects of GH and IGF-1 *in vivo.*

Function, Parameter Group	Function, Parameter Subgroup	GH	IGF-1
Carbohydrate metabolism	Glucose uptake in extrahepatic tissues	Decrease[a]	Increase
	Hepatic glucose output	Increase	Decrease
	Hepatic glycogen stores	Increase (jointly with glucocorticoids and insulin)	
	Plasma glucose	Increase	Decrease
	Insulin sensitivity	Decrease	Increase
Lipid metabolism	Lipolysis in adipocytes, plasma-free fatty acid levels	Increase	Decrease
	Plasma ketone bodies	Increase	Decrease
Protein metabolism (muscle, connective tissue)	Amino acid uptake	Increase (?)	Increase
	Protein synthesis	Increase (?)	Increase
	Nitrogen excretion	Decrease (?)	Decrease

[a]In GH-deficient patients, administration of GH results in a short-lived insulin-like action. During this time, glucose uptake by peripheral (extrahepatic) tissues increases.

Modified with permission from Kacsoh B. *Endocrine Physiology*. New York: McGraw-Hill Education; 2000.

TABLE 4–6   Main biologic effect of the GH-IGF-1 axis.

Target, Source	Parameter	Effect
Blood and plasma (liver, bone, and bone marrow actions)	IGF-1, acid-labile subunit IGF-binding protein-3 Alkaline phosphatase (bone specific) Fibrinogen Hemoglobin, hematocrit	Increased by GH only Increased by both GH and IGF-1 Increase (mainly IGF-1) Increase Increase (mainly IGF-1 action on bone marrow)
Cartilage, bone	Length (before epiphysial closure), width (periosteal and perichondrial growth)	Stimulation (mainly IGF-1)
Visceral organs (liver, spleen, thymus, thyroid), tongue and heart	Growth	Stimulation, organomegaly (both GH and IGF-1)
Renal 25-hydroxyvitamin D 1α-hydroxylase activity	Plasma calcitriol	Increase (mainly GH), promotes positive calcium balance
Kidney	GFR	Increase (IGF-1)
Skin	Hair growth Sweat glands Dermis	Stimulation (IGF-1?) Hyperplasia, hypertrophy, hyperfunction (GH?) Thickening (both GH and IGF-1)

Modified with permission from Kacsoh B. *Endocrine Physiology*. New York: McGraw-Hill Education; 2000.

proteolytic cleavage) and a low-affinity species. Measurement of serum concentrations of the high-affinity GHBP provides an index of GH receptor concentrations. For example, individuals with Laron dwarfism, a form of GH insensitivity characterized by mutations in the GH receptor, usually have abnormally low levels of GHBP.

Concentrations of IGF-1 are determined by radioreceptor assays or radioimmunoassays. Determining the levels of these mediators of GH action may result in more accurate assessment of the biologic activity of GH (see Chapter 6).

Secretion

The secretion of GH is predominantly mediated by two hypothalamic hormones: GHRH and somatostatin (GHIH), both of which contribute to the episodic pattern of GH secretion. These hypothalamic influences are tightly regulated by an integrated system of neural, metabolic, and hormonal factors. Table 4–7 summarizes the many factors that affect GH secretion in physiologic, pharmacologic, and pathologic states.

A. Growth hormone–releasing hormone   GHRH binds to specific receptors, stimulating cAMP production by somatotrophs and stimulating both GH synthesis and secretion. The effects of GHRH are partially blocked by somatostatin. The administration of GHRH to normal humans leads to rapid release of GH (within minutes); levels peak at 30 minutes and are sustained for 60 to 120 minutes.

Other peptide hormones such as ADH, ACTH, and α-MSH may act as GH-releasing factors when present in pharmacologic amounts. Even TSH and GnRH often cause GH secretion in patients with acromegaly; however, it is not certain whether any of these effects are mediated by the hypothalamus or represent direct effects on the somatotroph. Regulation of GHRH is primarily

under neural control (see discussion later), but there is also short-loop negative feedback by GHRH itself.

B. Somatostatin   Somatostatin, a tetradecapeptide, is a potent inhibitor of GH secretion. It decreases cAMP production in GH-secreting cells and inhibits both basal and stimulated GH secretion. Somatostatin secretion is increased by elevated levels of GH and IGF-1. Long-acting analogs of somatostatin have been used therapeutically in the management of GH excess and in conditions such as pancreatic and carcinoid tumors that cause diarrhea.

C. Growth hormone secretagogues   Non-GHRH secretagogues act to release GH, not through the GHRH receptor, but through a separate receptor, the growth hormone secretagogue receptor (GHS-R). A number of synthetic secretagogues, both peptides and nonpeptides, have been described. **Ghrelin**, a circulating peptide made by endocrine cells in the stomach, is the endogenous ligand for GHS-R (now termed ghrelin receptor). Its location in the stomach suggests a new mechanism for regulation of GH secretion.

D. Neural control   The neural control of basal GH secretion results in irregular and intermittent release associated with sleep and varying with age. Peak levels occur 1 to 4 hours after the onset of sleep (during stages 3 and 4) (Figure 4–14). These nocturnal sleep bursts, which account for nearly 70% of daily GH secretion, are greater in children and tend to decrease with age. Glucose infusion does not suppress this episodic release. Emotional, physical, and chemical stress, including surgery, trauma, exercise, electroshock therapy, and pyrogen administration, provoke GH release. In addition, impairment of secretion leading to growth failure has been well documented in children with severe emotional deprivation (see Chapter 6).

TABLE 4–7 Factors affecting GH secretion.

Increase	Decrease[a]
Physiologic	Postprandial hyperglycemia
Sleep	Elevated free fatty acids
Exercise	
Stress (physical or psychologic)	
Postprandial	
Hyperaminoacidemia	
Hypoglycemia (relative)	
Pharmacologic	Hormones
Hypoglycemia	Somatostatin
Absolute: insulin or 2-deoxyglucose	GH
Relative: postglucagon	Progesterone
Hormones	Glucocorticoids
GHRH	Neurotransmitters, etc
Ghrelin	Alpha-adrenergic
Peptide (ACTH, α-MSH, vasopressin)	antagonists
Estrogen	(phentolamine)
Neurotransmitters, etc	Beta-adrenergic agonists
Alpha-adrenergic agonists	(isoproterenol)
(clonidine)	Serotonin agonists
Beta-adrenergic antagonists	(methysergide)
(propranolol)	Dopamine antagonists
Serotonin precursors	(phenothiazines)
Dopamine agonists (levodopa,	
apomorphine, bromocriptine)	
GABA agonists (muscimol)	
Potassium infusion	
Pyrogens (pseudomonas	
endotoxin)	
Pathologic	Obesity
Protein depletion and starvation	Acromegaly; dopamine
Anorexia nervosa	agonists
Ectopic production of GHRH	Hypothyroidism
Chronic renal failure	Hyperthyroidism
Acromegaly	
TRH	
GnRH	

[a]Suppressive effects of some factors can be demonstrated only in the presence of a stimulus.

Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism*. 3rd ed. New York: McGraw-Hill Education; 1995.

E. Metabolic control   The metabolic factors affecting GH secretion include all fuel substrates: carbohydrate, protein, and fat. Glucose administration, orally or intravenously, lowers GH in healthy subjects and provides a simple physiologic maneuver useful in the diagnosis of acromegaly (see discussed later). In contrast, hypoglycemia stimulates GH release. This effect depends on intracellular glycopenia, because the administration of 2-deoxyglucose (a glucose analog that causes intracellular glucose deficiency) also increases GH. This response to hypoglycemia depends on both the rate of change in blood glucose and the absolute level attained.

A protein meal or intravenous infusion of amino acids (eg, arginine) causes GH release. Paradoxically, states of protein-calorie malnutrition also increase GH, possibly as a result of decreased IGF-1 production and lack of inhibitory feedback.

Fatty acids suppress GH responses to certain stimuli, including arginine and hypoglycemia. Fasting stimulates GH secretion, pos-

sibly as a means of mobilizing fat as an energy source and preventing protein loss.

F. Effects of other hormones   Responses to stimuli are blunted in states of cortisol excess and during hypo- and hyperthyroidism. Estrogen enhances GH secretion in response to stimulation.

G. Effects of neuropharmacologic agents   Many neurotransmitters and neuropharmacologic agents affect GH secretion. Biogenic amine agonists and antagonists act at the hypothalamic level and alter GHRH or somatostatin release. Dopaminergic, alpha-adrenergic, and serotonergic agents all stimulate GH release.

Dopamine agonists such as levodopa, apomorphine, and bromocriptine increase GH secretion, whereas dopaminergic antagonists such as phenothiazines inhibit GH. The effect of levodopa, a precursor of both norepinephrine and dopamine, may be mediated by its conversion to norepinephrine, because its effect is blocked by the alpha-adrenergic antagonist phentolamine. Moreover, phentolamine suppresses GH release in response to other stimuli such as hypoglycemia, exercise, and arginine, emphasizing the importance of alpha-adrenergic mechanisms in modulating GH secretion.

Beta-adrenergic agonists inhibit GH, and beta-adrenergic antagonists such as propranolol enhance secretion in response to provocative stimuli.

PROLACTIN

Biosynthesis

PRL is a 198-amino-acid polypeptide hormone (MW 22,000) synthesized and secreted from the lactotrophs of the anterior pituitary. Despite evolution from an ancestral hormone common to GH and human placental lactogen (hPL), PRL shares only 16% of its residues with the former and 13% with hPL. A precursor molecule (MW 40,000-50,000) is also secreted and may constitute 8% to 20% of the PRL plasma immunoreactivity in healthy persons and in patients with PRL-secreting pituitary tumors. PRL and GH are structurally related to members of the cytokine-hematopoietin family that include erythropoietin, granulocyte-macrophage colony stimulating factor (GM-CSF), and interleukins IL-2 to IL-7.

Function

PRL stimulates lactation in the postpartum period (see Chapter 16). During pregnancy, PRL secretion increases and, in concert with many other hormones (estrogen, progesterone, hPL, insulin, and cortisol), promotes additional breast development in preparation for milk production. Despite its importance during pregnancy, PRL has not been demonstrated to play a role in the development of normal breast tissue in humans. During pregnancy, estrogen enhances breast development but blunts the effect of PRL on lactation; the decrease in both estrogen and progesterone after parturition allows initiation of lactation. Accordingly, galactorrhea
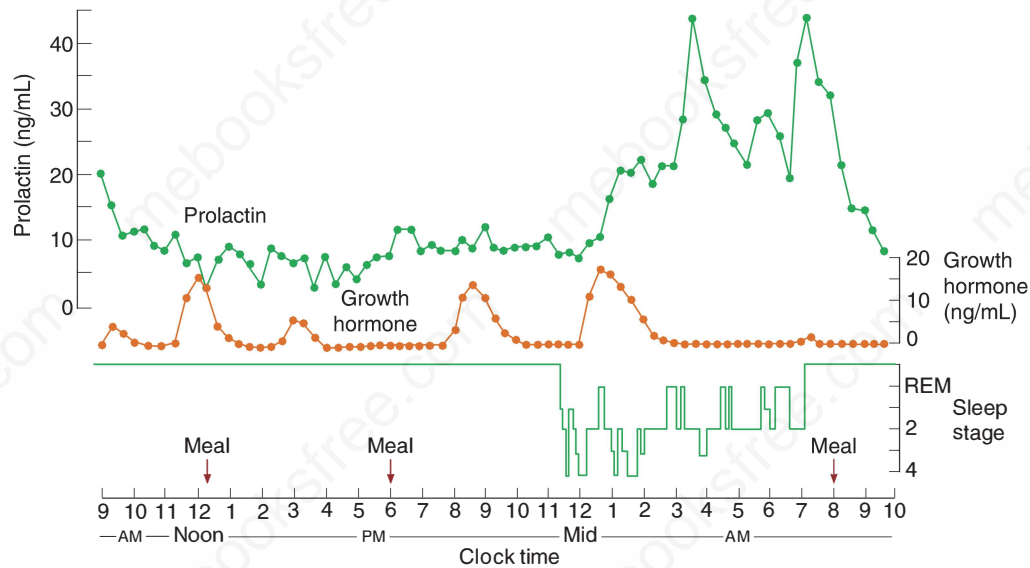
FIGURE 4–14 Sleep-associated changes in prolactin (PRL) and growth hormone (GH) secretion in humans. Peak levels of GH occur during sleep stages 3 or 4; the increase in PRL is observed 1 to 2 hours after sleep begins and is not associated with a specific sleep phase. (Reproduced with permission from Sassin JF, Frantz AG, Weitzman ED, et al. Human prolactin: 24-hour pattern with increased release during sleep. *Science*. 1972 Sep 29;177(4055):1205-1207.)

may accompany the discontinuance of oral contraceptives or estrogen therapy. Although basal PRL secretion falls in the postpartum period, lactation is maintained by persistent breast suckling.

PRL levels are very high in the fetus and in newborn infants, declining during the first few months of life.

Although PRL does not appear to play a physiologic role in the regulation of gonadal function, hyperprolactinemia in humans leads to hypogonadism. In women, initially there is a shortening of the luteal phase; subsequently, anovulation, oligomenorrhea or amenorrhea, and infertility occur. In men, PRL excess leads to decreased testosterone synthesis and decreased spermatogenesis, which clinically present as decreased libido, impotence, and infertility. The exact mechanisms of PRL inhibition of gonadal function are unclear, but the principal one appears to be alteration of hypothalamic-pituitary control of gonadotropin secretion. Basal LH and FSH levels are usually normal; however, their pulsatile secretion is decreased and the midcycle LH surge is suppressed in women. Gonadotropin reserve, as assessed by administration of exogenous GnRH, is usually normal. PRL also has a role in immunomodulation; extrapituitary synthesis of PRL occurs in T lymphocytes (among other sites), and PRL receptors are present on T and B lymphocytes and macrophages. PRL modulates and stimulates both immune cell proliferation and survival.

Measurement

The PRL secretory rate is approximately 400 μg/d (18.6 nmol/d). The hormone is cleared by the liver (75%) and the kidney (25%), and its half-time of disappearance from plasma is about 25 to 50 minutes.

Basal levels of PRL in adults vary considerably, with a mean of 13 ng/mL (0.6 nmol/L) in women and 5 ng/mL (0.23 nmol/L)

in men. The upper range of normal in most laboratories is 15 to 20 ng/mL (0.7-0.9 nmol/L).

PRL is measured using a highly specific immunoradiometric assay. However, when PRL levels are extremely high, which can occur with some PRL-secreting tumors, this assay may be susceptible to the "hook effect." PRL levels are erroneously reported as normal or modestly elevated due to saturation of available assay antibodies. Appropriate sample dilutions (ie, 1:100) will avoid this artifact.

In some patients a form of PRL with molecular mass greater than 150 kDa predominates. This is termed *macroprolactinemia* and consists of aggregates of monomeric PRL as well as PRL-immunoglobulin G complexes. These complexes may have reduced biological activity and can be measured using polyethylene glycol precipitation of serum samples.

Secretion

The hypothalamic control of PRL secretion is predominantly inhibitory, and dopamine is the most important inhibitory factor. The physiologic, pathologic, and pharmacologic factors influencing PRL secretion are listed in Table 4–8.

A. Prolactin-releasing factors   TRH is a potent PRL-releasing factor that evokes release of PRL at a threshold dose similar to that which stimulates release of TSH. An exaggerated response of both TSH and PRL to TRH is observed in primary hypothyroidism, and their responses are blunted in hyperthyroidism. In addition, PRL secretion is also stimulated by VIP and serotonergic pathways.

B. Episodic and sleep-related secretion   PRL secretion is episodic. An increase is observed 60 to 90 minutes after sleep

TABLE 4–8 Factors affecting prolactin secretion.

Increase	Decrease
Physiologic Pregnancy Nursing Nipple stimulation Exercise Stress (hypoglycemia) Sleep Seizures Neonatal	

Pharmacologic
TRH
Estrogen
Vasoactive intestinal peptide
Dopamine antagonists
  (phenothiazines, haloperidol,
  risperidone, metoclopramide,
  reserpine, methyldopa,
  amoxapine, opioids)
Monoamine oxidase inhibitors
Verapamil
Licorice
	Dopamine agonists
	  (levodopa, apomorphine,
	  bromocriptine, pergolide)
	GABA

Pathologic Pituitary tumors Hypothalamic/pituitary stalk lesions Neuraxis irradiation Chest wall lesions Spinal cord lesions Hypothyroidism Chronic renal failure Severe liver disease	Pseudohypoparathyroidism Pituitary destruction or removal Lymphocytic hypophysitis

begins but—in contrast to GH—is not associated with a specific sleep phase. Peak levels are usually attained between 4 and 7 AM (see Figure 4–14). This sleep-associated augmentation of PRL release is not part of a circadian rhythm, like that of ACTH; it is related strictly to the sleeping period regardless of when it occurs during the day.

C. Other stimuli   Stresses, including surgery, exercise, hypoglycemia, and acute myocardial infarction, cause significant elevation of PRL levels. Nipple stimulation in nonpregnant women also increases PRL. This neurogenic reflex may also occur from chest wall injury such as mechanical trauma, burns, surgery, and herpes zoster of thoracic dermatomes. This reflex discharge of PRL is abolished by denervation of the nipple or by spinal cord or brain stem lesions.

D. Effects of other hormones   Many hormones influence PRL release. Estrogens augment basal and stimulated PRL secretion after 2 to 3 days of use (an effect that is of special clinical importance in patients with PRL-secreting pituitary adenomas); glucocorticoids tend to suppress TRH-induced PRL secretion; and thyroid hormone administration may blunt the PRL response to TRH.

E. Effects of pharmacologic agents (Table 4–8)   Many pharmacologic agents alter PRL secretion. Dopamine agonists (eg,

bromocriptine) decrease secretion, forming the basis for their use in states of PRL excess. Dopamine antagonists (eg, receptor blockers such as phenothiazines and metoclopramide) and dopamine depletors (eg, reserpine) augment PRL release. Serotonin agonists enhance PRL secretion; serotonin receptor blockers suppress PRL release associated with stress and with nursing.

THYROTROPIN

Biosynthesis

TSH is a glycoprotein (MW 28,000) composed of two noncovalently linked alpha and beta subunits. The structure of the alpha subunit of TSH is identical to that of the other glycoprotein molecules—FSH, LH, and hCG—but the beta subunits differ in these glycoproteins and is responsible for their biologic and immunologic specificity. The peptides of these subunits appear to be synthesized separately and united before the carbohydrate groups are attached. The intact molecule is then secreted, as are small amounts of nonlinked subunits.

Function

The beta subunit of TSH attaches to high-affinity receptors in the thyroid, stimulating iodide uptake, hormonogenesis, and release of T_4 and T_3. This occurs through activation of adenylyl cyclase and the generation of cAMP. TSH secretion also causes an increase in gland size and vascularity by promoting mRNA and protein synthesis. (For a more detailed description, see Chapter 7.)

Measurement

TSH circulates unbound in the blood with a half-life of 35 to 50 minutes. With ultrasensitive immunometric assays for measuring TSH concentration, the normal range is usually 0.5 to 4.7 μU/mL (0.5-4.7 mU/L). Measurement of TSH is the most sensitive screening test for the diagnosis of primary hypothyroidism and hyperthyroidism; however, TSH levels alone cannot be used to evaluate pituitary or hypothalamic hypothyroidism.

The alpha subunit can be detected in about 80% of normals, with a range of 0.5 to 2 ng/mL. Plasma alpha subunit levels increase after administration of TRH in normal subjects, and basal levels are elevated in primary hypothyroidism, primary hypogonadism, and in patients with TSH-secreting, gonadotropin-secreting, or pure alpha subunit-secreting pituitary adenomas.

Secretion

The secretion of TSH is controlled by both stimulatory (TRH) and inhibitory (somatostatin) influences from the hypothalamus and in addition is modulated by the feedback inhibition of thyroid hormone on the hypothalamic-pituitary axis.

A. TRH   The response of TSH to TRH is modulated by the circulating concentration of thyroid hormones. Small changes in serum levels (even within the physiologic range) cause substantial alterations in the TSH response to TRH. As shown in
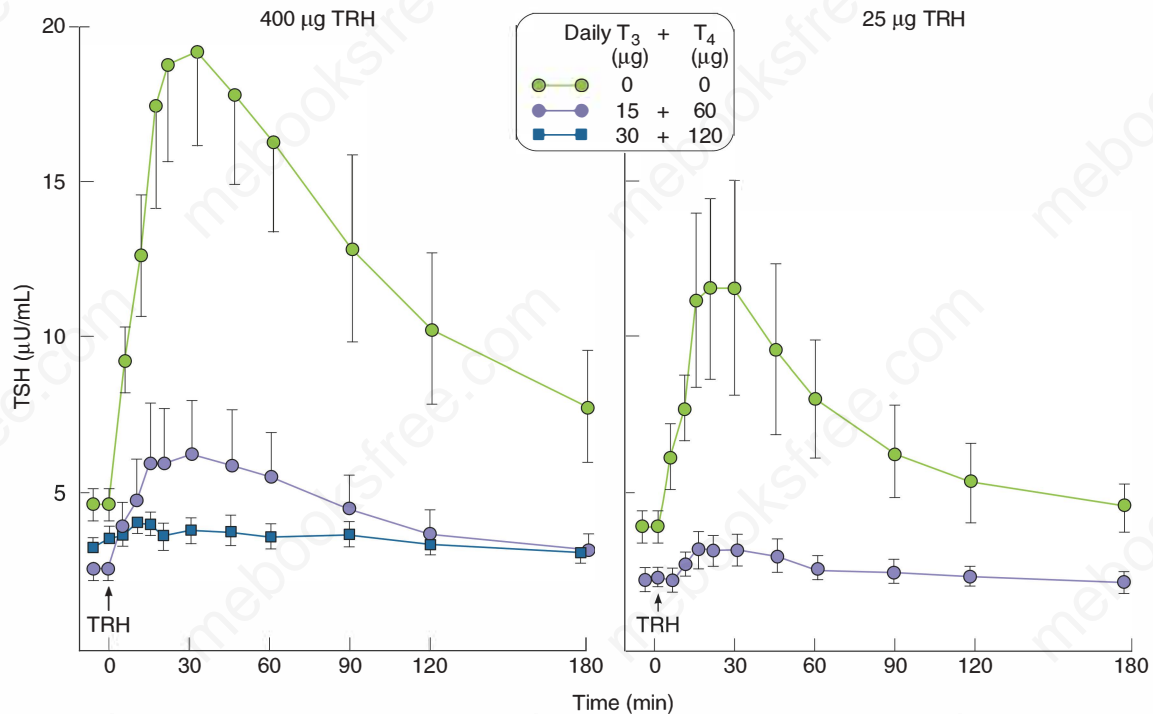
FIGURE 4–15 Administration of small doses of T_3 (15 µg) and T_4 (60 µg) to healthy subjects inhibits the TSH response to two doses (400 µg, left; 25 µg, right) of TRH (protirelin). (Reproduced with permission from Snyder PJ, Utiger RD. Inhibition of thyrotropin-releasing hormone by small quantities of thyroid hormones. *J Clin Invest.* 1972 Aug;51(8):2077-2084.)

Figure 4–15, the administration of T_3 (15 µg) and T_4 (60 µg) to healthy persons for 3 to 4 weeks suppresses the TSH response to TRH despite only small increases in circulating T_3 and T_4 levels. Thus, the secretion of TSH is inversely proportional to the concentration of thyroid hormone.

The set point (the level at which TSH secretion is maintained) is determined by TRH. Deviations from this set point result in appropriate changes in TSH release. Administration of TRH increases TSH within 2 minutes, and this response is blocked by previous T_3 administration; however, larger doses of TRH may overcome this blockade—suggesting that both T_3 and TRH act at the pituitary level to influence TSH secretion. In addition, T_3 and T_4 inhibit mRNA for TRH synthesis in the hypothalamus, indicating that a negative feedback mechanism operates at this level also.

B. Somatostatin This inhibitory hypothalamic peptide augments the direct inhibitory effect of thyroid hormone on the thyrotrophs. Infusion of somatostatin blunts the early morning TSH surge and suppresses high levels of TSH in primary hypothyroidism. Octreotide acetate, a somatostatin analog, has been used successfully to inhibit TSH secretion in patients with TSH-secreting pituitary tumors.

C. Neural control In addition to these hypothalamic influences on TSH secretion, neurally mediated factors may be important. Dopamine physiologically inhibits TSH secretion. Intravenous dopamine administration decreases TSH in both healthy and hypothyroid subjects as well as blunts the TSH response to TRH. Thus, as expected, dopaminergic agonists such as bromocriptine

inhibit TSH secretion and dopaminergic antagonists such as metoclopramide increase TSH secretion in euthyroid subjects.

D. Effects of cortisol and estrogens Glucocorticoid excess has been shown to impair the sensitivity of the pituitary to TRH and may lower serum TSH to undetectable levels. However, estrogens increase the sensitivity of the thyrotroph to TRH; women have a greater TSH response to TRH than men do, and pretreatment of men with estradiol increases their TRH-induced TSH response. (See also Chapter 7.)

E. Other Serum TSH levels exhibit a diurnal variation with the peak occurring around midnight and the nadir at 8 to 10 AM. TSH levels are also higher in older individuals and those with higher body mass indices although the relationship is poorly understood (TSH decreases in patients who lose weight after bariatric surgery). Some authorities advocate for a higher TSH goal in elderly patients (TSH of 4 to 6 µU/mL for age >70 years).

GONADOTROPINS: LUTEINIZING HORMONE (LH) AND FOLLICLE-STIMULATING HORMONE (FSH)

Biosynthesis

LH and FSH are glycoprotein gonadotropins composed of alpha and beta subunits and secreted by the same cell. The specific beta subunit confers on these hormones their unique biologic activity,

as it does with TSH and hCG. The biologic activity of hCG, a placental glycoprotein, closely resembles that of LH. Human menopausal gonadotropin (hMG, menotropins)—an altered mixture of pituitary gonadotropins recovered from the urine of postmenopausal women—is a preparation with FSH-like activity. Menotropins and hCG are used clinically for induction of spermatogenesis or ovulation (see Chapters 12 and 13).

Function

LH and FSH bind to receptors in the ovary and testis and regulate gonadal function by promoting sex steroid production and gametogenesis.

In men, LH stimulates testosterone production from the interstitial cells of the testes (Leydig cells). Maturation of spermatozoa, however, requires both LH and FSH. FSH stimulates testicular growth and enhances the production of an androgen-binding protein by the Sertoli cells, which are a component of the testicular tubule necessary for sustaining the maturing sperm cell. This androgen-binding protein promotes high local concentrations of testosterone within the testis, an essential factor in the development of normal spermatogenesis (see Chapter 12).

In women, LH stimulates estrogen and progesterone production from the ovary. A surge of LH in the mid-menstrual cycle is responsible for ovulation, and continued LH secretion subsequently stimulates the corpus luteum to produce progesterone by enhancing the conversion of cholesterol to pregnenolone. Development of the ovarian follicle is largely under FSH control, and the secretion of estrogen from this follicle is dependent on both FSH and LH.

Measurement

The normal levels of LH and FSH vary with the age of the subject (see Appendix). They are low before puberty and elevated in postmenopausal women. A nocturnal rise of LH in boys and the cyclic secretion of FSH and LH in girls usually herald the onset of puberty before clinical signs are apparent. In women, LH and FSH vary during the menstrual cycle; during the initial phase of the cycle (follicular), LH steadily increases, with a midcycle surge that initiates ovulation. FSH, on the other hand, initially rises and then decreases during the later follicular phase until the midcycle surge, which is concordant with LH. Both LH and FSH levels fall steadily after ovulation (see Chapter 13).

LH and FSH levels in men are similar to those in women during the follicular phase. The alpha subunit, shared by all the pituitary glycoprotein hormones, can also be measured (see TSH) and will rise following GnRH administration.

Secretion

The secretion of LH and FSH is controlled by GnRH, which maintains basal gonadotropin secretion, generates the phasic release of gonadotropins for ovulation, and determines the onset of puberty. As noted earlier, the size and frequency of GnRH pulses determine the ratio of gonadotropin secretion: low-frequency pulses favor FSH release while high-frequency pulses result in LH release. Many other factors are involved in regulation of this axis. For example, activins and follistatins are paracrine factors that exert opposing effects on gonadotrophs. Leptin, a hormone made in adipocytes in proportion to fat stores, is involved in regulation of this axis and may help to explain the suppression of gonadotropin secretion that accompanies caloric restriction.

A. Episodic secretion   In both males and females, secretion of LH and FSH is episodic, with secretory bursts that occur each hour and are mediated by a concordant episodic release of GnRH. The amplitude of these secretory surges is greater in patients with primary hypogonadism. The pulsatile nature of GnRH release is critical for sustaining gonadotropin secretion. A continuous, prolonged infusion of GnRH in women evokes an initial increase in LH and FSH followed by prolonged suppression of gonadotropin secretion. This phenomenon may be explained by downregulation of GnRH receptors on the pituitary gonadotrophs. Consequently, long-acting synthetic analogs of GnRH may be used clinically to suppress LH and FSH secretion in conditions such as precocious puberty.

B. Positive feedback   Circulating sex steroids affect GnRH secretion and thus LH and FSH secretion by both positive and negative (inhibitory) feedback mechanisms. During the menstrual cycle, estrogens provide a positive influence on GnRH effects on LH and FSH secretion, and the rise in estrogen during the follicular phase is the stimulus for the LH and FSH ovulatory surge. This phenomenon suggests that the secretion of estrogen is to some extent influenced by an intrinsic ovarian cycle. Progesterone amplifies the duration of the LH and FSH surge and augments the effect of estrogen. After this midcycle surge, the developed egg leaves the ovary. Ovulation occurs approximately 10 to 12 hours after the LH peak and 24 to 36 hours after the estradiol peak. The remaining follicular cells in the ovary are converted, under the influence of LH, to a progesterone-secreting structure, the corpus luteum. After about 12 days, the corpus luteum involutes, resulting in decreased estrogen and progesterone levels and then uterine bleeding (see Chapter 13).

C. Negative feedback   Negative feedback effects of sex steroids on gonadotropin secretion also occur. In women, primary gonadal failure or menopause results in elevations of LH and FSH, which can be suppressed with long-term, high-dose estrogen therapy. However, a shorter duration of low-dose estrogen may enhance the LH response to GnRH. In men, primary gonadal failure with low circulating testosterone levels is also associated with elevated gonadotropins. However, testosterone is not the sole inhibitor of gonadotropin secretion in men, since selective destruction of the tubules (eg, by cyclophosphamide therapy) results in azoospermia and elevation of only FSH.

Inhibin, a polypeptide (MW 32,000) secreted by the Sertoli cells of the seminiferous tubules, is the major factor that inhibits FSH secretion by negative feedback. Inhibin consists of separate alpha and beta subunits connected by a disulfide bridge. Androgens stimulate inhibin production; this peptide may help to locally regulate spermatogenesis (see Chapter 12).

ENDOCRINOLOGIC EVALUATION OF THE HYPOTHALAMIC-PITUITARY AXIS

The precise assessment of the hypothalamic-pituitary axis has been made possible by immunoassays of the major anterior pituitary hormones and their specific target gland hormones. In addition, provocative testing using synthetic or purified hormones (eg, ACTH, ovine CRH, glucagon, and insulin) can be used to assess hypothalamic-pituitary reserve and excess.

This section describes the principles involved in testing each pituitary hormone as well as special situations (eg, drugs, obesity) that may interfere with pituitary function or pituitary testing. Specific protocols for performing and interpreting diagnostic procedures are outlined at the end of this section and in Table 4–9. The clinical manifestations of either hypo- or hypersecretion of anterior pituitary hormones are discussed in subsequent sections.

EVALUATION OF ADRENOCORTICOTROPIC HORMONE

ACTH deficiency leads to secondary adrenocortical insufficiency, characterized by decreased secretion of cortisol and the adrenal androgens; aldosterone secretion, controlled primarily by the renin-angiotensin axis, is usually maintained. In contrast, excessive ACTH secretion leads to adrenal hyperfunction (Cushing syndrome, discussed in a later section of this chapter and in Chapter 9).

Plasma ACTH Levels

Basal ACTH measurements are usually unreliable indicators of pituitary function, because its short plasma half-life and episodic secretion result in wide fluctuations in plasma levels (see Figure 4–13). Therefore, the interpretation of plasma ACTH levels requires the simultaneous assessment of cortisol secretion by the adrenal cortex. These measurements are of greatest utility in differentiating primary and secondary adrenocortical insufficiency and in establishing the etiology of Cushing syndrome (see the later section on Cushing disease and also Chapter 9).

Evaluation of ACTH Deficiency

In evaluating ACTH deficiency, measurement of basal cortisol levels is also generally unreliable. Because plasma cortisol levels are usually low in the late afternoon and evening, reflecting the normal diurnal rhythm, samples drawn at these times are of virtually no value for this diagnosis. Plasma cortisol levels are usually highest in the early morning; however, there is considerable overlap between adrenal insufficiency and normal subjects. A plasma cortisol level less than 5 μg/dL (138 nmol/L) at 8 AM strongly suggests the diagnosis—and the lower the level, the more likely the diagnosis. Conversely, a plasma cortisol greater than 20 μg/dL (552 nmol/L) virtually excludes the diagnosis. Similarly, salivary cortisol levels less than 1.8 ng/mL (5 nmol/L) at 8 AM strongly suggest adrenal insufficiency, while levels in excess of 5.8 ng/mL

(16 nmol/L) greatly reduce the probability of the diagnosis. Consequently, the diagnosis of ACTH hyposecretion (secondary adrenal insufficiency) must be established by provocative testing of the reserve capacity of the hypothalamic-pituitary axis.

Adrenal Stimulation

Because adrenal atrophy develops as a consequence of prolonged ACTH deficiency, the initial and most convenient approach to evaluation of the hypothalamic-pituitary-adrenal axis is assessment of the plasma cortisol response to synthetic ACTH (cosyntropin). In normal individuals, injection of cosyntropin (250 μg) causes a rapid increase (within 30 minutes) of cortisol to at least 18 to 20 μg/dL (496-552 nmol/L), and this response usually correlates with the cortisol response to insulin-induced hypoglycemia (typically considered the gold standard in the diagnosis of adrenocortical insufficiency). With the introduction of more specific cortisol assays (eg, tandem mass spectrometry), the normal peak cortisol response following ACTH stimulation may be lower than currently accepted ranges. A subnormal cortisol response to ACTH confirms adrenocortical insufficiency. However, a normal response does not directly evaluate the ability of the hypothalamic-pituitary axis to respond to stress (see Chapter 9). Thus, patients withdrawn from long-term glucocorticoid therapy may have an adequate increase in cortisol following exogenous ACTH that precedes complete recovery of the hypothalamic-pituitary-adrenal axis. Therefore, such patients should receive glucocorticoids during periods of stress for at least 1 year after steroids are discontinued, unless the hypothalamic-pituitary axis is shown to be responsive to stress as described later.

The more physiologic dose administered in the 1 μg ACTH test is designed to improve its sensitivity in detection of secondary adrenal insufficiency. The cortisol response to 1 μg of synthetic ACTH correlates better with the cortisol response to insulin-induced hypoglycemia in patients with complete chronic secondary adrenal insufficiency. However, the results in partial secondary adrenal insufficiency are less reliable and technical difficulties make the test impractical for routine use. Thus, the standard 250 μg ACTH test remains the procedure of choice (see discussion later and Chapter 9).

Pituitary Stimulation

Direct evaluation of pituitary ACTH reserve can be performed by means of insulin-induced hypoglycemia, metyrapone administration, or CRH stimulation. These studies are unnecessary if the cortisol response to rapid ACTH stimulation is subnormal.

A. Insulin-induced hypoglycemia The stimulus of neuroglycopenia associated with hypoglycemia (blood glucose <40 mg/dL) evokes a stress-mediated activation of the hypothalamic-pituitary-adrenal axis. Subjects should experience adrenergic symptoms (diaphoresis, tachycardia, weakness, headache) associated with the fall in blood sugar. In normal persons, plasma cortisol increases to more than 18 to 20 μg/dL (496-552 nmol/L), indicating normal ACTH reserve. Although plasma ACTH also rises, its

TABLE 4–9 Endocrine tests of hypothalamic-pituitary function.

	Method	Sample Collection	Possible Side Effects; Contraindications	Interpretation
Rapid ACTH stimulation test (cosyntropin test)	Administer synthetic ACTH (1-24) (cosyntropin), 250 µg intravenously or intramuscularly. The test may be performed at any time of the day or night and does not require fasting. The low-dose test is performed in the same manner except that 1 µg of synthetic ACTH (1-24) is administered.	Obtain samples for plasma cortisol at 0 and 30 min or at 0 and 60 min.	Rare allergic reactions have been reported.	A normal response is a peak plasma cortisol level >18-20 µg/dL (496-552 nmol/L).
Insulin hypoglycemia test	Give nothing by mouth after midnight. Start an intravenous infusion with normal saline solution. Regular insulin is given intravenously in a dose sufficient to cause adequate hypoglycemia (blood glucose <40 mg/dL). The dose is 0.1-0.15 unit/kg (healthy subjects); 0.2-0.3 unit/kg (obese subjects or those with Cushing syndrome or acromegaly); 0.05 unit/kg (patients with suspected hypopituitarism).	Collect blood for glucose determinations every 15 min during the study. Samples of GH and cortisol are obtained at 0, 30, 45, 60, 75, and 90 min.	A physician must be in attendance. Symptomatic hypoglycemia (diaphoresis, headache, tachycardia, weakness) is necessary for adequate stimulation and occurs 20-35 min after insulin is administered in most patients. If severe central nervous system signs or symptoms occur, intravenous glucose (25-50 mL of 50% glucose) should be given immediately; otherwise, the test can be terminated with a meal or with oral glucose. This test is contraindicated in the elderly or in patients with cardiovascular or cerebrovascular disease and seizure disorders.	Symptomatic hypoglycemia and a fall in blood glucose to <40 mg/dL (2.2 mmol/L) will increase GH to a maximal level >5 ng/mL; some investigators regard an increment of 6 ng/mL (280 pmol/L) as normal. Plasma cortisol should increase to a peak level of at least 18-20 µg/dL (496-552 nmol/L).
Metyrapone test	Metyrapone is given orally between 11 and 12 PM with a snack to minimize gastrointestinal discomfort. The dose is 30 mg/kg.	Blood for plasma 11-deoxycortisol and cortisol determinations is obtained at 8 AM the morning after metyrapone is given.	Gastrointestinal upset may occur. Adrenal insufficiency may occur. Metyrapone should not be used in sick patients or those in whom primary adrenal insufficiency is suspected.	Serum 11-deoxycortisol should increase to >7 µg/dL (190 nmol/L). Cortisol should be <10 µg/dL (280 nmol/L) in order to ensure adequate inhibition of 11β-hydroxylation.
Glucagon stimulation test	The patient should be fasting after midnight. Give glucagon 1 mg intramuscularly.	Blood for plasma GH and capillary blood glucose at 0, 30, 60, 90, 120, 150, and 180 min.	Nausea and late hypoglycemia. This test is contraindicated in malnourished patients or patients who have not eaten for >48 h.	GH rises to >3 ng/mL in normal individuals. Glucose usually rises to peak around 90 min, then gradually declines.
Glucose growth hormone suppression test	The patient should be fasting after midnight; give glucose, 75-100 g orally.	GH and glucose should be determined at 0, 30, and 60 min after glucose administration.	Patients may complain of nausea after the large glucose load.	GH levels are suppressed to <2 ng/mL (90 pmol/L) in healthy subjects. Failure of adequate suppression or a paradoxic rise may be seen in acromegaly, starvation, protein-calorie malnutrition, and anorexia nervosa.
Clomiphene test	Clomiphene is administered orally. For women, give 100 mg daily for 5 d (beginning on day 5 of the cycle if the patient is menstruating); for men, give 100 mg daily for 7-10 d.	Blood for LH and FSH determinations is drawn before and after clomiphene is given.	This drug may stimulate ovulation, and women should be advised accordingly.	In women, LH and FSH levels peak on the fifth day to a level above the normal range. After the fifth day, LH and FSH levels decline. In men, LH should double after 1 wk; FSH will also increase, but to a lesser extent.

(Continued)

TABLE 4–9 Endocrine tests of hypothalamic-pituitary function. (*Continued*)

	Method	Sample Collection	Possible Side Effects; Contraindications	Interpretation
CRH test	CRH (1 µg/kg) is given intravenously as a bolus injection.	Blood samples for ACTH and cortisol are taken at 0, 15, 30, and 60 min.	Flushing often occurs. Transient tachycardia and hypotension have also been reported.	The ACTH response is dependent on the assay utilized and occurs 15 min after CRH is administered. The peak cortisol response occurs at 30-60 min and is usually >10 µg/dL (276 nmol/L).
Low-dose dexamethasone suppression test	Dexamethasone (1 mg) is given between 11 PM and midnight.	Blood samples for cortisol and dexamethasone are taken at 8 AM the next morning.	Side effects are rare. Compliance is sometimes an issue. Some medications and patient variability can affect dexamethasone metabolism.	Cortisol should suppress to <1.8 µg/dL (50 nmol/L) in normal individuals. This cutoff has a high sensitivity, but specificity is compromised.

determination has not proved to be as useful, because pulsatile secretion requires frequent sampling, and the normal response is not well standardized. Although insulin-induced hypoglycemia most reliably predicts ACTH secretory capacity in times of stress, it is rarely performed at present, because the procedure requires a physician's presence and is contraindicated in elderly patients, patients with cerebrovascular or cardiovascular disease, and those with seizure disorders. It should be used with caution in patients in whom diminished adrenal reserve is suspected, because severe hypoglycemia may occur; in these patients, the test should always be preceded by the ACTH adrenal stimulation test.

B. Metyrapone stimulation   Metyrapone administration is an alternative method for assessing ACTH secretory reserve. Metyrapone inhibits P450c11 (11β-hydroxylase), the enzyme that catalyzes the final step in cortisol biosynthesis (see Chapter 9). The inhibition of cortisol secretion interrupts negative feedback on the hypothalamic-pituitary axis, resulting in a compensatory increase in ACTH. The increase in ACTH secretion stimulates increased steroid biosynthesis proximal to P450c11, and the increase can be detected as an increase in the precursor steroid (11-deoxycortisol) in plasma. The overnight test is preferred because of its simplicity; it is performed by administering 30 mg/kg of metyrapone orally at midnight. Plasma 11-deoxycortisol is determined the following morning and rises to more than 7 µg/dL (0.2 nmol/L) in healthy individuals. Again, the test should be used cautiously in patients with suspected adrenal insufficiency and should be preceded by a rapid ACTH stimulation test (see discussion earlier). The traditional 3-day metyrapone test should not be used at present because of the risk of precipitating adrenal insufficiency. The overnight metyrapone test is most useful in patients with partial secondary adrenal insufficiency in whom the rapid ACTH stimulation test is normal or borderline. It has been shown to correlate well with the response to insulin-induced hypoglycemia. .

C. CRH stimulation   Ovine CRH administered intravenously is used to assess ACTH secretory dynamics. In healthy subjects, CRH (1 µg/kg) provokes a peak ACTH response within 15 minutes and a peak cortisol response within 30 to 60 minutes. This dose may be associated with mild flushing, occasional shortness of breath, tachycardia, and hypotension. Patients with primary adrenal insufficiency have elevated basal ACTH levels and exaggerated ACTH responses to CRH. Secondary adrenal insufficiency results in an absent ACTH response to CRH in patients with pituitary corticotroph destruction; however, in patients with hypothalamic dysfunction, there is a prolonged and augmented ACTH response to CRH with a delayed peak. Since there are few data available on the ACTH and cortisol responses to CRH in normal subjects compared to those with secondary adrenal insufficiency, this study may have limitations.

ACTH Hypersecretion

ACTH hypersecretion is manifested by adrenocortical hyperfunction (Cushing syndrome). The diagnosis and differential diagnosis of ACTH hypersecretion and hypercortisolism are outlined in a later section on Cushing disease and also in Chapter 9. The low-dose dexamethasone suppression test is used to establish the presence of hypercortisolism regardless of its cause. Dexamethasone 1 mg is taken orally between 11 PM and midnight. The following morning at 8 AM, a blood sample is obtained for cortisol (and dexamethasone level if available or cortisol does not suppress). Cortisol values less than 1.8 µg/dL are normal. Sensitivity using this cutoff is good though specificity is compromised. Patient compliance, individual differences in dexamethasone metabolism, drugs that affect metabolism of dexamethasone, and medical conditions that alter cortisol secretion (acute illness, anxiety, depression, alcoholism) may influence the accuracy of this test.

EVALUATION OF GROWTH HORMONE

The evaluation of GH secretory reserve is important in the assessment of children with short stature and in adults with suspected hypopituitarism. Provocative tests are necessary because basal levels of GH are usually low and, therefore, do not distinguish between normal and GH-deficient patients. Special attention must be given to the methodology and the laboratory standards of GH measurement. Newer immunoassays give results that are 30% to 50% lower than older radioimmunoassays.

Insulin-Induced Hypoglycemia

The most reliable stimulus of GH secretion is insulin-induced hypoglycemia. In normal individuals, GH levels increase to more than 5 ng/mL after adequate hypoglycemia is achieved. Because 10% of normal individuals fail to respond to hypoglycemia, other stimulatory tests may be necessary.

GHRH-Arginine Test

Both forms of human GHRH (GHRH-40 and GHRH-44) have been used to evaluate GH secretory capacity. A dose of GHRH (1 µg/kg) combined with a 30-minute infusion of arginine (0.5 g/kg to a maximum of 20 g), which presumably reduces hypothalamic somatostatin secretion, promptly stimulates GH; the mean peak is 10 to 15 ng/mL (460-700 pmol/L) at 30 to 60 minutes in healthy subjects. The results are comparable to those achieved with insulin-induced hypoglycemia. Unfortunately, GHRH is no longer available in the United States.

Glucagon Stimulation Test

Glucagon can also be used to evaluate GH secretory capacity. A dose of glucagon (1 mg) is given intramuscularly, and GH levels are measured every 30 minutes for 3 to 4 hours. In adults with GH deficiency, peak GH levels fail to rise above 3 ng/mL.

Tests with Levodopa, Arginine, and Other Stimuli

Stimulation testing with levodopa, arginine infusion alone, or propranolol, is less reliable in the diagnosis of GH deficiency.

GH Hypersecretion

The evaluation of GH hypersecretion is discussed in the section on acromegaly and is most conveniently assessed by GH suppression testing with oral glucose and measurement of IGF-1 levels.

EVALUATION OF PROLACTIN

PRL secretion by the pituitary is resistant to local damage; decreased PRL secretory reserve indicates severe intrinsic pituitary disease.

PRL hypersecretion is a common endocrine problem. Macroprolactinemia can be assessed using polyethylene glycol precipitation of serum samples. Additional evaluation of hyperprolactinemia is discussed in the section on prolactinomas.

EVALUATION OF THYROID-STIMULATING HORMONE

Basal Measurements

The laboratory evaluation of TSH secretory reserve begins with an assessment of target gland secretion; thyroid function tests (free

thyroxine [FT_4]) should be obtained. Normal thyroid function studies in a clinically euthyroid patient indicate adequate TSH secretion, and no further studies are warranted. Laboratory evidence of hypothyroidism requires measurement of a TSH level. With primary thyroid gland failure, the TSH level is elevated; low or normal TSH in the presence of hypothyroidism suggests hypothalamic-pituitary dysfunction (see Chapter 7).

TRH Test

Because accurate methods for determining TSH and FT_4 readily establish the diagnosis of hypothyroidism in virtually all patients, the TRH test is rarely indicated today (TRH [protirelin] is also not currently commercially available).

EVALUATION OF LH AND FSH

Testosterone and Estrogen Levels

The evaluation of gonadotropin function also requires assessment of target gland secretory function, and measurement of gonadal steroids (testosterone in men, estradiol in women) is useful in the diagnosis of hypogonadism. In women, the presence of regular menstrual cycles is strong evidence that the hypothalamic-pituitary-gonadal axis is intact. Estradiol levels rarely fall below 50 pg/mL (180 pmol/L), even during the early follicular phase. A level of less than 30 pg/mL (110 pmol/L) in the presence of oligomenorrhea or amenorrhea is indicative of gonadal failure. In men, serum testosterone (normal range, 300-1000 ng/dL; 10-35 nmol/L) is a sensitive index of gonadal function (see Chapters 12 and 13).

LH and FSH Levels

In the presence of gonadal insufficiency, high LH and FSH levels are a sign of primary gonadal failure; low or normal LH and FSH suggest hypothalamic-pituitary dysfunction (hypogonadotropic hypogonadism).

GnRH Test

LH and FSH secretory reserves may be assessed with the use of synthetic GnRH (gonadorelin). Administration of GnRH causes a prompt increase in plasma LH and a lesser and slower increase in FSH. However, in most patients, the GnRH test provides no more useful information than is obtained by measurement of basal gonadotropin and gonadal steroid levels. Thus, this test is uncommonly performed. (GnRH is also not commercially available in the United States.)

PROBLEMS IN EVALUATION OF THE HYPOTHALAMIC-PITUITARY AXIS

This section briefly outlines some of the disorders and conditions that may cause confusion and lead to misinterpretation of pituitary function tests. The effects of drugs are described in the next section.

Obesity

GH dynamics are impaired in many severely obese patients; all provocative stimuli, including insulin-induced hypoglycemia, arginine, levodopa, and glucagon plus propranolol, often fail to provoke GH secretion. The GH response to GHRH is also impaired in obesity and improves with weight loss. Obesity is also a common cause of hypogonadotropic hypogonadism in adult men.

Diabetes Mellitus

Although glucose normally suppresses GH secretion, most individuals with type 1 diabetes have normal or elevated GH levels that often do not rise further in response to hypoglycemia or arginine. Levodopa increases GH in some diabetic patients, and even a dopamine infusion (which produces no GH change in nondiabetic subjects, because it does not cross the blood–brain barrier), stimulates GH in diabetic patients. Despite the increased GH secretion in patients with inadequately controlled diabetes, the GH response to GHRH in insulin-dependent diabetic patients is similar to that of nondiabetic subjects. IGF-1 levels are low in insulin-deficient diabetes despite the elevated GH levels.

Uremia

Basal levels of GH, PRL, LH, FSH, TSH, and free cortisol tend to be elevated, for the most part owing to prolongation of their plasma half-life. GH may paradoxically increase following glucose administration and is often hyperresponsive to a hypoglycemic stimulus. Although the administration of TRH has no effect on GH secretion in healthy subjects, the drug may increase GH in patients with chronic renal failure. The response of PRL to TRH is blunted and prolonged. Gonadotropin response to synthetic GnRH usually remains intact. Dexamethasone suppression of cortisol may be impaired.

Starvation and Anorexia Nervosa

GH secretion increases with fasting and malnutrition, and such conditions may cause a paradoxical increase in GH following glucose administration. Severe starvation, such as occurs in patients with anorexia nervosa, may result in low levels of gonadal steroids. LH and FSH responses to GnRH may be intact despite a state of functional hypogonadotropic hypogonadism. Cortisol levels are frequently increased and fail to suppress adequately with dexamethasone. PRL and TSH dynamics are usually normal despite a marked decrease in circulating total thyroid hormones (see Chapter 7).

Depression

Depression may alter the ability of dexamethasone to suppress plasma cortisol and may elevate cortisol secretion; the response to insulin-induced hypoglycemia usually remains intact. In addition, late-evening salivary cortisol levels usually remain normal and are not elevated as seen in patients with Cushing syndrome. The ACTH response to CRH is blunted in endogenous depression. Some depressed patients also have abnormal GH dynamics: TRH may increase GH, and hypoglycemia or levodopa may fail to increase GH. These patients may also show blunted TSH responses to TRH.

Pharmacologic Agents and Alcohol

Glucocorticoid excess impairs the GH response to hypoglycemia, the TSH response to TRH, and the LH response to GnRH. Estrogens tend to augment GH dynamics as well as the PRL and TSH response to TRH. Estrogens increase plasma cortisol secondary to a rise in corticosteroid-binding globulin and may result in inadequate suppression with dexamethasone.

Phenytoin enhances the metabolism of dexamethasone, making clinical testing in patients of this agent difficult to interpret. Phenothiazines may blunt the GH response to hypoglycemia and levodopa and frequently cause hyperprolactinemia. The many other pharmacologic agents that increase PRL secretion are listed in Table 4–8.

Narcotics, including heroin, morphine, and methadone, may all raise PRL levels and suppress GH and cortisol response to hypoglycemia. Opioids will also suppress hypothalamic CRH and thereby decrease ACTH and cortisol secretion. Hypogonadism is common with narcotic use.

In chronic alcoholics, alcohol excess or withdrawal may increase cortisol levels and cause inadequate dexamethasone suppression and an impaired cortisol increase after hypoglycemia.

ENDOCRINE TESTS OF HYPOTHALAMIC-PITUITARY FUNCTION

Protocols for performing endocrine tests and their normal responses are summarized in Table 4–9. The indications for and the clinical utility of these procedures are described in the preceding section and will be mentioned again in the section on pituitary and hypothalamic disorders.

NEURORADIOLOGIC EVALUATION

Symptoms of pituitary hormone excess or deficiency, headache, or visual disturbance lead the clinician to consider a hypothalamic-pituitary disorder. In this setting, accurate neuroradiologic assessment of the hypothalamus and pituitary is essential in confirming the existence and defining the extent of hypothalamic-pituitary lesions. However, the diagnosis of such lesions should be based on both endocrine and radiologic criteria, because variability of pituitary anatomy in the normal population may lead to false-positive interpretations. Furthermore, patients with pituitary microadenomas may have normal neuroradiologic studies. Imaging studies must be interpreted in light of the fact that 10% to 20% of the general population harbor nonfunctional and asymptomatic pituitary microadenomas.

Magnetic Resonance Imaging (MRI)

MRI is the current procedure of choice for imaging the hypothalamus and pituitary. It has superseded the use of computed tomography (CT) because it allows better definition of normal structures and has better resolution in defining tumors. Arteriography is rarely utilized at present except in patients with intrasellar or parasellar aneurysms.

Imaging is performed in sagittal and coronal planes at 1.5 to 2 mm intervals. This allows clear definition of hypothalamic and pituitary anatomy and accurately visualize lesions as small as 3 to 5 mm. The use of the heavy-metal contrast agent gadolinium allows even more precise differentiation of small pituitary adenomas from normal anterior pituitary tissue and other adjacent structures as shown in Figure 4–16.

A. Normal anatomy The normal anterior pituitary is 5 to 7 mm in height and approximately 10 mm in its lateral dimensions. The superior margin is flat or concave but may be upwardly convex with a height of 10 to 12 mm in healthy menstruating young women. The floor of the sella turcica is formed by the bony roof of the sphenoid sinus, and its lateral margins are formed by the dural membranes of the cavernous sinuses, which contain the carotid arteries and the third, fourth, and sixth cranial nerves. The posterior pituitary appears on MRI as a high-signal-intensity structure, the "posterior pituitary bright spot," which is absent in patients with diabetes insipidus. The pituitary stalk, which is normally in the midline, is 2 to 3 mm in diameter and 5 to 7 mm in length. The pituitary stalk joins the inferior hypothalamus below the third ventricle and posterior to the optic chiasm. All of these normal structures are readily visualized with MRI; the normal pituitary
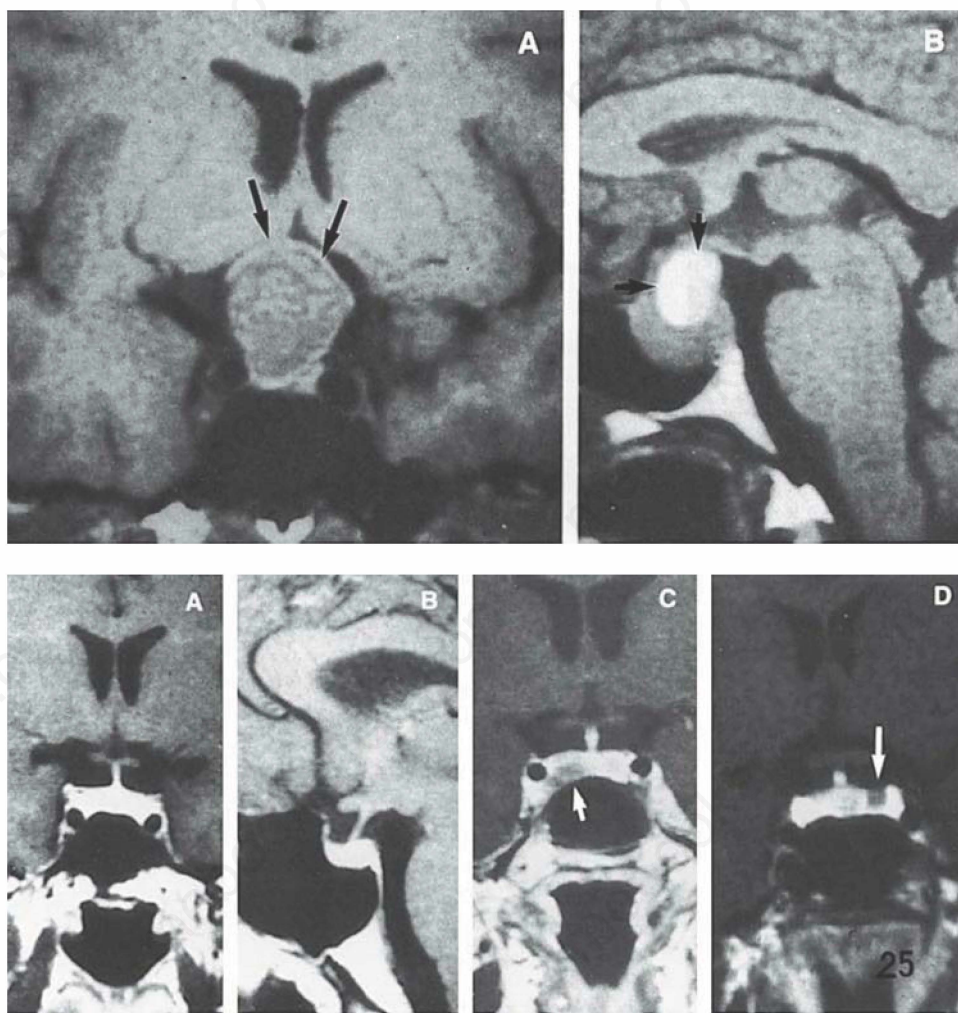
FIGURE 4–16 **Upper panel: A**: The coronal magnetic resonance (MR) image shows a large nonfunctioning pituitary adenoma (arrows) with pronounced suprasellar extension and chiasmal compression. **B**: A sagittal MR image of another large pituitary adenoma shows spontaneous hemorrhage within the suprasellar portion of the adenoma (arrows). **Lower panel**: Gadolinium-enhanced MR images are shown of the pituitary gland. **A and B**: Coronal and sagittal images show the normal, uniformly enhancing pituitary stalk and pituitary gland. **C**: A pituitary microadenoma appears as a low-intensity lesion in the inferior aspect of the right lobe of the gland (arrow). **D**: The pituitary microadenoma appears as a low-intensity lesion between the left lobe of the pituitary and the left cavernous sinus (arrow). (Used with permission from David Norman, MD.)

and the pituitary stalk show increased signal intensity with gadolinium.

B. Microadenomas   These lesions, which range from 2 to 10 mm in diameter, appear as low-signal-intensity lesions on MRI and do not usually enhance with gadolinium. Adenomas less than 5 mm in diameter may not be visualized and do not usually alter the normal pituitary contour. Lesions greater than 5 mm in diameter create a unilateral convex superior gland margin and usually cause deviation of the pituitary stalk toward the side opposite the adenoma.

MRI scans must be interpreted with caution, because minor abnormalities occur in 10% of patients who have had incidental high-resolution scans but no clinical pituitary disease. These abnormalities may represent the clinically insignificant pituitary abnormalities that occur in 10% to 20% of the general population, and they may also be due to small intrapituitary cysts, which usually occur in the pars intermedia. Artifacts within the sella turcica associated with the bones of the skull base may also result in misinterpretation of imaging studies. Finally, many patients with pituitary microadenomas have normal high-resolution MRI scans. Therefore, despite increased accuracy of neuroradiologic diagnosis, the presence or absence of a small pituitary tumor and the decision concerning its treatment must be based on the entire clinical picture.

C. Macroadenomas   Pituitary adenomas greater than 10 mm in diameter are readily visualized with MRI scans, and the scan also defines the adjacent structures and degree of extension of the lesion. Thus, larger tumors show compression of the normal pituitary and distortion of the pituitary stalk. Adenomas larger than 1.5 cm frequently have suprasellar extension, and MRI scans show compression and upward displacement of the optic chiasm. Less commonly, there is lateral extension and invasion of the cavernous sinus.

D. Other uses   High-resolution MRI scanning is also a valuable tool in the diagnosis of empty sella syndrome, hypothalamic tumors, and other parasellar lesions.

PITUITARY AND HYPOTHALAMIC DISORDERS

Hypothalamic-pituitary lesions present with a variety of manifestations, including pituitary hormone hypersecretion and hyposecretion, sellar enlargement, headache, and visual loss. The approach to evaluation should be designed to ensure early diagnosis at a stage when the lesions are amenable to therapy.

Etiology and Early Manifestations

In adults, the most common cause of hypothalamic-pituitary dysfunction is a pituitary adenoma. Some of the earliest symptoms of pituitary tumors are due to endocrinologic abnormalities. These usually precede sellar enlargement and local manifestations such as

headache and visual loss which are late manifestations seen in patients with larger tumors or suprasellar extension.

In children, pituitary adenomas are uncommon; the most frequent structural lesions causing hypothalamic-pituitary dysfunction are craniopharyngiomas and other hypothalamic tumors. These also usually manifest as endocrine disturbances (low GH levels, delayed puberty, diabetes insipidus) prior to the development of headache, visual loss, or other central nervous system symptoms.

Common and Later Manifestations

A. Pituitary hypersecretion   PRL is the hormone most commonly secreted in excess amounts by pituitary adenomas, and it is usually elevated in patients with hypothalamic disorders and pituitary stalk compression as well. Thus, PRL measurement is essential in evaluating patients with suspected pituitary disorders and should be performed in patients presenting with galactorrhea, gonadal dysfunction, secondary gonadotropin deficiency, or enlargement of the sella turcica. Hypersecretion of GH or ACTH leads to the more characteristic syndromes of acromegaly and Cushing disease, respectively (see discussion later).

B. Pituitary insufficiency   Although panhypopituitarism is a classic manifestation of pituitary adenomas, it is present in less than 20% of patients in current large series because of earlier diagnosis of these lesions.

The earliest clinical manifestation of a pituitary adenoma in adults is hypogonadism secondary to elevated levels of PRL, GH, or ACTH and cortisol. The hypogonadism in these patients is due to interference with the secretion of GnRH rather than to destruction of anterior pituitary tissue. Thus, patients with hypogonadism should first be screened with FSH and LH measurements to exclude primary gonadal failure (elevated FSH or LH) and those with hypogonadotropic hypogonadism should have serum PRL levels measured and be examined for clinical evidence of GH or ACTH and cortisol excess.

In children, short stature is the most frequent clinical presentation of hypothalamic-pituitary dysfunction; in these patients, GH deficiency should be considered (see Chapter 6).

TSH or ACTH deficiency is relatively unusual and, if present, usually indicates panhypopituitarism. Thus, patients with secondary hypothyroidism or hypoadrenalism should undergo a complete assessment of pituitary function and neuroradiologic studies, because panhypopituitarism and large pituitary tumors are common in this setting. PRL measurement is again essential, because prolactinomas are the most frequent pituitary tumors in adults.

C. Enlarged sella turcica   Patients may present with enlargement of the sella turcica, which may be noted on radiographs performed for head trauma or on sinus series. These patients usually have either a pituitary adenoma or empty sella syndrome. Other less common causes include craniopharyngioma, lymphocytic hypophysitis, and carotid artery aneurysm. Evaluation should include clinical assessment of pituitary dysfunction and measurements of PRL and thyroid and adrenal function. Pituitary

function is usually normal in the empty sella syndrome; this diagnosis can be confirmed by MRI. Patients with clinical or laboratory evidence of pituitary dysfunction usually have a pituitary adenoma.

D. Visual field defects   Patients presenting with bitemporal hemianopsia or unexplained visual field defects or visual loss should be considered to have a pituitary or hypothalamic disorder until proven otherwise. The initial steps in diagnosis should be neuro-ophthalmologic evaluation and neuroradiologic studies with MRI, which will reveal the tumor if one is present. These patients should also have PRL measurements and be assessed for anterior pituitary insufficiency, which is especially common with large pituitary adenomas.

In addition to causing visual field defects, large pituitary lesions may extend laterally into the cavernous sinus, compromising the function of the third, fourth, or sixth cranial nerve, leading to diplopia.

E. Diabetes insipidus   Diabetes insipidus is a common manifestation of hypothalamic lesions and metastases to the pituitary, but simply never occurs in primary pituitary lesions. Diagnostic evaluation is described in Chapter 5. In addition, all patients should undergo radiologic evaluation and assessment of anterior pituitary function.

EMPTY SELLA SYNDROME

Etiology and Incidence

Empty sella syndrome occurs when the subarachnoid space extends into the sella turcica, partially filling it with cerebrospinal fluid. This process causes remodeling and enlargement of the sella turcica and flattening of the pituitary gland.

Primary empty sella syndrome resulting from congenital incompetence of the diaphragma sellae (Figure 4–17) is common, with an incidence in autopsy series ranging from 5% to 23%. It is the most frequent cause of enlarged sella turcica. An empty sella is also commonly seen after pituitary surgery or radiation therapy and may also occur following postpartum pituitary infarction (Sheehan syndrome). In addition, both PRL-secreting and GH-secreting pituitary adenomas may undergo subclinical hemorrhagic infarction and cause contraction of the overlying suprasellar cistern downward into the sella. Therefore, the presence of an empty sella does not exclude the possibility of a coexisting pituitary tumor.

Clinical Features

A. Symptoms and signs   Most patients are middle-aged obese women. Many have systemic hypertension; benign intracranial hypertension may also occur. Although 48% of patients complain of headache, this feature may have only initiated the evaluation (ie, imaging of the brain/head), and its relationship to the empty sella is probably coincidental. Serious clinical manifestations are uncommon. Spontaneous cerebrospinal fluid rhinorrhea and visual field impairment may occur rarely.

B. Laboratory findings   Tests of anterior pituitary function are almost always normal, although some patients have hyperprolactinemia. Endocrine function studies should be performed to exclude pituitary hormone insufficiency or a hypersecretory pituitary microadenoma.

Diagnosis

The diagnosis of empty sella syndrome can be readily confirmed by MRI, which demonstrates the herniation of the diaphragma sellae and the presence of cerebrospinal fluid in the sella turcica.
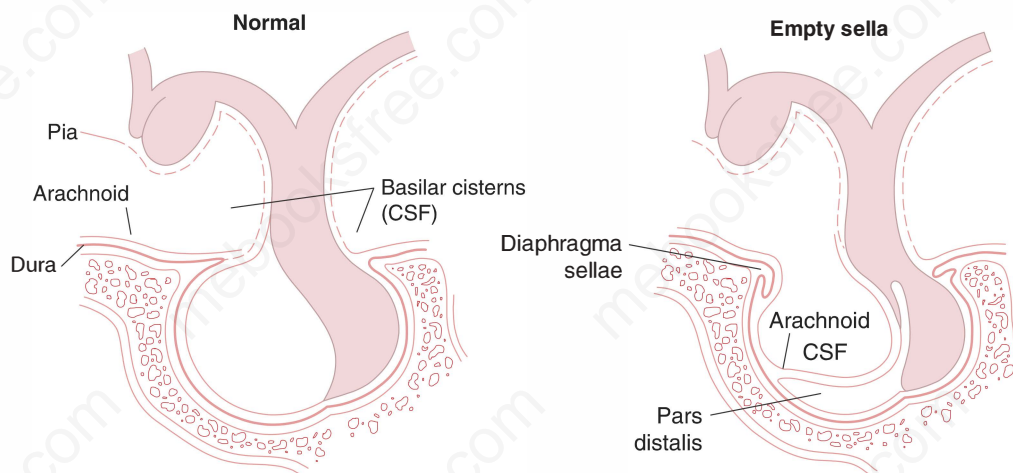
FIGURE 4–17 Representation of the normal relationship of the meninges to the pituitary gland **(left)** and the findings in the empty sella **(right)** as the arachnoid membrane herniates through the incompetent diaphragma sellae. (Reproduced with permission from Jordan RM, Kendall JW, Kerber CW. The primary empty sella syndrome: analysis of the clinical characteristics, radiographic features, pituitary function, and cerebrospinal fluid adenohypophysial hormone concentrations. *Am J Med*. 1977 Apr;62(4):569-580.)

HYPOTHALAMIC DYSFUNCTION

Hypothalamic dysfunction is most often caused by tumors, of which craniopharyngioma is the most common in children, adolescents, and young adults. In older adults, primary central nervous system tumors and those arising from hypothalamic (epidermoid and dermoid tumors) and pineal structures (pinealomas) are more common. Other causes of hypothalamic-pituitary dysfunction are discussed later in the section on hypopituitarism.

Clinical Features

A. Craniopharyngioma   Craniopharyngiomas are thought to originate from metaplasia of remnant epithelial cell rests in Rathke pouch and the craniopharyngeal duct during development. They are classified as adamantinomatous or papillary. Most adamantinomatous tumors have mutations in the beta-catenin gene while BRAF mutations have been found in papillary craniopharyngiomas. The differential diagnosis includes Rathke cleft cysts and arachnoid cysts. Rathke cleft cysts are a common incidental finding within the pituitary, but may present with symptoms similar to craniopharyngiomas. Although Rathke cleft cysts and craniopharyngiomas are both thought to arise from Rathke pouch during development, craniopharyngiomas tend to be more aggressive and are more likely to recur after resection.

The initial symptoms of craniopharyngioma in children and adolescents are predominantly endocrinologic; however, these manifestations are frequently unrecognized, and at diagnosis more than 80% of patients have hypothalamic-pituitary endocrine deficiencies. These endocrine abnormalities may precede presenting symptoms by months or years; GH deficiency is most common, with about 50% of patients having growth retardation and approximately 70% decreased GH responses to stimulation at diagnosis. Gonadotropin deficiency leading to absent or arrested puberty is usual in older children and adolescents; TSH and ACTH deficiencies are less common, and diabetes insipidus is present in about 15%.

Symptoms leading to the diagnosis are frequently neurologic and due to the mass effect of the expanding tumor. Symptoms of increased intracranial pressure such as headache and vomiting are present in about 40% of patients; decreased visual acuity or visual field defects are the presenting symptoms in another 35%. MRI confirms the tumor in virtually all patients; in 95%, the tumor is suprasellar.

In adults, craniopharyngiomas have similar presentations; that is, the diagnosis is usually reached as a result of investigation of complaints of headache or visual loss. However, endocrine manifestations—especially hypogonadism, diabetes insipidus, or other deficiencies of anterior pituitary hormones—usually precede these late manifestations. MRI readily demonstrates the tumors, which in adults are almost always both intrasellar and suprasellar. The appearance is typically of a heterogeneous, cystic mass that enhances with contrast. Calcifications may also be present.

B. Other tumors   Other hypothalamic or pineal tumors and primary central nervous system tumors involving the hypothalamus have variable presentations in both children and adults. Thus, presentation is with headache, visual loss, symptoms of increased intracranial pressure, growth failure, various degrees of hypopituitarism, or diabetes insipidus. Endocrine deficiencies usually precede neurologic manifestations. Hypothalamic tumors in childhood may present with precocious puberty.

C. Other manifestations of hypothalamic dysfunction
Lesions in the hypothalamus can cause many other abnormalities, including disorders of consciousness, behavior, thirst, appetite, and temperature regulation. These abnormalities are usually accompanied by hypopituitarism and diabetes insipidus.

Somnolence can occur with hypothalamic lesions, as can a variety of changes in emotional behavior. Decreased or absent thirst may occur and predispose these patients to dehydration. When diminished thirst accompanies diabetes insipidus, fluid balance is difficult to control. Hypothalamic dysfunction may also cause increased thirst, leading to polydipsia and polyuria that may mimic diabetes insipidus. Obesity is common in patients with hypothalamic tumors because of hyperphagia, decreased satiety, and decreased activity. Frohlich syndrome, or adiposogenital dystrophy, is characterized by obesity, growth retardation, and delayed development of sexual organs. Anorexia and weight loss are unusual manifestations of these tumors.

Temperature regulation can also be disordered in these patients. Sustained or, less commonly, paroxysmal hyperthermia can occur following acute injury due to trauma, hemorrhage, or craniotomy. This problem usually lasts less than 2 weeks. Poikilothermia, the inability to adjust to changes in ambient temperature, can occur in patients with bilateral hypothalamic lesions. These patients most frequently exhibit hypothermia but can also develop hyperthermia during hot weather. A few patients manifest sustained hypothermia due to anterior hypothalamic lesions.

Diagnosis

Patients with suspected hypothalamic tumors should undergo MRI to determine the extent and nature of the tumor. Complete assessment of anterior pituitary function is necessary in these patients, because deficiencies are present in the great majority (see section on Hypopituitarism later), and the evaluation will establish the requirements for replacement therapy. PRL levels should also be determined, because most hypothalamic lesions cause hyperprolactinemia either by hypothalamic injury or by damage to the pituitary stalk.

Treatment

Treatment depends on the type of tumor. Complete resection of craniopharyngioma is usually not feasible, and neurologic complications can be devastating with aggressive intervention. Therefore, this tumor is best managed by observation and replacement of insufficient hormones, or limited neurosurgical removal of accessible tumor and decompression of cysts, possibly followed by radiotherapy. Patients treated by this method have a recurrence rate of approximately 20%; with surgery alone, the recurrence rate approximates 80%.

Other hypothalamic tumors are usually not completely resectable; however, biopsy is indicated to arrive at a histologic diagnosis.

HYPOPITUITARISM

Hypopituitarism is manifested by diminished or absent secretion of one or more pituitary hormones. The development of signs and symptoms is often slow and insidious, depending on the rate of onset and the magnitude of hypothalamic-pituitary damage—factors that are influenced by the underlying pathogenesis. Hypopituitarism is either a primary event caused by destruction of the anterior pituitary gland or a secondary phenomenon resulting from deficiency of hypothalamic stimulatory factors normally acting on the pituitary. In general, acquired loss of anterior pituitary function follows the sequence of GH, LH/FSH, TSH, ACTH, and PRL. Treatment and prognosis depend on the extent of hypofunction, the underlying cause, and the location of the lesion in the hypothalamic-pituitary axis.

Etiology

The etiologic considerations in hypopituitarism are diverse. As shown later, a helpful mnemonic device is the phrase "nine Is": invasive, infarction, infiltrative, injury, immunologic, iatrogenic, infectious, idiopathic, and isolated. Most of these lesions may cause pituitary or hypothalamic failure (or both). Establishing the precise cause of hypopituitarism is helpful in determining treatment and prognosis.

A. Invasive   Space-occupying lesions cause hypopituitarism by destroying the pituitary gland or hypothalamic nuclei or by disrupting the hypothalamic-hypophysial portal venous system. Large pituitary adenomas cause hypopituitarism by these mechanisms, and pituitary function may improve after their removal. Small pituitary tumors—microadenomas (<10 mm in diameter)—characteristically seen in the hypersecretory states (excess PRL, GH, ACTH) do not directly cause pituitary insufficiency. Craniopharyngioma, the most common tumor of the hypothalamic-pituitary region in children, frequently impairs pituitary function by its compressive effects. Primary central nervous system tumors, including meningioma, chordoma, optic glioma, epidermoid tumors, and dermoid tumors, may decrease hypothalamic-pituitary secretion by their mass effects. Metastatic lesions to this area are common (especially breast carcinoma) but rarely result in clinically obvious hypopituitarism. Anatomic malformations such as basal encephalocele and parasellar aneurysms cause hypothalamic-pituitary dysfunction and may enlarge the sella turcica and mimic pituitary tumors.

B. Infarction   Ischemic damage to the pituitary has long been recognized as a cause of hypopituitarism. In 1914, Simmonds reported pituitary necrosis in a woman with severe puerperal sepsis. In 1937 Sheehan published his classic description of its occurrence following postpartum hemorrhage and vascular collapse.

The mechanism for the ischemia in such cases is not certain. Hypotension along with vasospasm of the hypophysial arteries is currently believed to compromise arterial perfusion of the anterior pituitary. During pregnancy, the pituitary gland may be more sensitive to hypoxemia because of its increased metabolic needs or more susceptible to the prothrombotic effects of the hyperestrogenic state. Some investigators have noted that the hypopituitarism does not always correlate with the degree of hemorrhage but that there is good correlation between the pituitary lesion and severe disturbances of the clotting mechanism (as in patients with placenta previa). Ischemic pituitary necrosis has also been reported to occur with greater frequency in patients with diabetes mellitus.

The extent of pituitary damage determines the rapidity of onset as well as the magnitude of pituitary hypofunction. The gland has a great secretory reserve, and more than 75% must be destroyed before clinical manifestations are evident. The initial clinical feature in postpartum necrosis may be failure to lactate after parturition; failure to resume normal menstrual periods is another clue to the diagnosis. However, the clinical features of hypopituitarism are often subtle, and years may pass before pituitary insufficiency is recognized following an ischemic insult.

Spontaneous hemorrhagic infarction of a pituitary tumor (pituitary apoplexy) frequently results in partial or total pituitary insufficiency. Pituitary apoplexy is often a fulminant clinical syndrome manifested by severe headache, visual impairment, ophthalmoplegias, meningismus, and an altered level of consciousness. Pituitary apoplexy is usually associated with a pituitary tumor; it may also be related to diabetes mellitus, radiotherapy, or open heart surgery. Acute pituitary failure with hypotension may result, and rapid mental deterioration, coma, and death may ensue. Emergency treatment with corticosteroids (see Chapter 24) and transsphenoidal decompression of the intrasellar contents may be lifesaving and may prevent permanent visual loss. Most patients who have survived pituitary apoplexy have developed multiple adenohypophysial deficits, but infarction of the tumor in some patients may cure the hypersecretory pituitary adenoma and its accompanying endocrinopathy. Pituitary infarction may also be a subclinical event (silent pituitary apoplexy), resulting in improvement of pituitary hormone hypersecretion without impairing the secretion of other anterior pituitary hormones.

C. Infiltrative   Hypopituitarism may be the initial clinical manifestation of infiltrative disease processes such as sarcoidosis, hemochromatosis, and Langerhan histiocytosis.

1. **Sarcoidosis**—The most common intracranial sites of involvement of sarcoidosis are the hypothalamus and pituitary gland. At one time, the most common endocrine abnormality in patients with sarcoidosis was thought to be diabetes insipidus; however, many of these patients actually have centrally mediated disordered control of thirst that results in polydipsia and polyuria, which in some cases explains the abnormal water metabolism. Deficiencies of multiple anterior pituitary hormones have been well documented in sarcoidosis and are usually secondary to hypothalamic insufficiency. Granulomatous involvement of the hypothalamic-pituitary unit is occasionally

extensive, resulting in visual impairment, and therefore, may simulate the clinical presentation of a pituitary or hypothalamic tumor.

2. **Hemochromatosis**—Hypopituitarism, particularly hypogonadotropic hypogonadism, is a prominent manifestation of iron storage disease—either primary (hereditary) hemochromatosis or transfusional iron overload. Hypogonadism occurs in most such cases and is often the initial clinical feature of iron excess; complete iron studies should be considered in any patient presenting with unexplained hypogonadotropic hypogonadism. If the diagnosis is established early, hypogonadism in hemochromatosis may be reversible with iron depletion. Pituitary deficiencies of TSH, GH, and ACTH may occur later in the course of the disease and are not reversible by iron chelation therapy.

3. **Langerhan histiocytosis**—In this disorder, the infiltration of multiple organs by well-differentiated histiocytes, is often heralded by the onset of diabetes insipidus and anterior pituitary hormone deficiencies. Most histologic and biochemical studies have indicated that this infiltrative process involves chiefly the hypothalamus, and hypopituitarism occurs only as a result of hypothalamic damage.

D. Injury   Severe head trauma may cause anterior pituitary insufficiency and diabetes insipidus. Posttraumatic anterior hypopituitarism may be due to injury to the anterior pituitary, the pituitary stalk, or the hypothalamus. Pituitary insufficiency with growth retardation has been described in battered children who suffer closed head trauma with subdural hematoma.

E. Immunologic   Lymphocytic hypophysitis resulting in anterior hypopituitarism is a distinct entity, occurring most often in women during pregnancy or in the postpartum period. It may present as a mass lesion of the sella turcica with visual field disturbances simulating pituitary adenoma. An autoimmune process with extensive infiltration of the gland by lymphocytes and plasma cells destroys the anterior pituitary cells. These morphologic features are similar to those of other autoimmune endocrinopathies (eg, thyroiditis, adrenalitis, and oophoritis). About 50% of patients with lymphocytic hypophysitis have other endocrine autoimmune disease, and circulating pituitary autoantibodies have been found in several cases. It is presently uncertain how this disorder should be diagnosed and treated. It must be considered in the differential diagnosis of women with pituitary gland enlargement and hypopituitarism during pregnancy or the postpartum period. Hypophysitis may also be a component of immunoglobulin G4 (IgG4)-related systemic syndrome. In addition to findings of sellar mass and/or thickened pituitary stalk on MRI, many patients have elevated circulating IgG4 levels. Pathology of the pituitary lesion and other affected organs is characterized by a lymphoplasmacytic infiltrate enriched in IgG4-positive plasma cells. Most patients respond to glucocorticoid treatments.

Lymphocytic hypophysitis may result in isolated hormone deficiencies (especially ACTH or prolactin). Consequently, women with this type of hypopituitarism may continue to menstruate while suffering from secondary hypothyroidism or hypoadrenalism.

A newer class of medications, checkpoint inhibitors, have been associated with hypophysitis. Ipilimumab, an anti-neoplastic monoclonal antibody used to treat melanoma and other malignancies, has been shown to cause hypophysitis in 11% to 17% of patients in one review. Median time to diagnosis after initiation was 8 to 10 weeks. Most patients have pituitary enlargement, but visual defects are rare. Although pituitary enlargement resolves, hypopituitarism remains in most patients and ACTH deficiency may be the only anterior pituitary hormone deficit.

F. Iatrogenic   Both surgical and radiation therapy to the pituitary gland may compromise its function. The anterior pituitary is quite resilient during transsphenoidal microsurgery, and despite extensive manipulation during the search for microadenomas, anterior pituitary function is usually preserved. The dose of conventional radiation therapy presently employed to treat pituitary tumors is 4500 to 5000 cGy and results in a 50% to 60% incidence of hypothalamic and pituitary insufficiency. Such patients most frequently have modest hyperprolactinemia (PRL 30-100 ng/mL [1.3-4.5 nmol/L]) with GH and gonadotropin failure; TSH and ACTH deficiencies are less common. Heavy particle (proton beam) irradiation and gamma knife radiosurgery for pituitary tumors result in a 15% to 55% incidence of hypopituitarism. Irradiation of tumors of the head and neck (nasopharyngeal cancer, brain tumors) and prophylactic cranial irradiation in leukemia may also cause hypopituitarism. The clinical onset of pituitary failure in such patients is usually insidious and results from both pituitary and hypothalamic injury.

G. Infectious   Although many infectious diseases, including tuberculosis, syphilis, and mycotic infections, have been implicated as causative agents in pituitary hypofunction, antimicrobial drugs have now made them rare causes of hypopituitarism.

H. Idiopathic   In some patients with hypopituitarism, no underlying cause is found. These may be isolated (see discussion later) or multiple deficiencies. Familial forms of hypopituitarism characterized by a small, normal, or enlarged sella turcica have been described. Both autosomal recessive and X-linked recessive inheritance patterns have been reported. A variety of complex congenital disorders may include deficiency of one or more pituitary hormones (eg, Prader-Willi syndrome, septo-optic dysplasia). Although progress has been made into understanding the genetic basis for some of these disorders, the pathogenesis remains uncertain.

I. Isolated   Isolated (monotropic) deficiencies of the anterior pituitary hormones have been described. Some of these have been associated with mutations in the genes coding for the specific hormones. Others, particularly GH deficiency, have been associated with mutations in genes necessary for normal pituitary development as noted.

1. **GH deficiency**—In children, congenital monotropic GH deficiency may be sporadic or familial. These children, who may experience fasting hypoglycemia, have a gradual deceleration in

growth velocity after 6 to 12 months of age. Diagnosis must be based on failure of GH responsiveness to provocative stimuli and the demonstration of normal responsiveness of other anterior pituitary hormones. Monotropic GH deficiency and growth retardation have also been observed in children suffering severe emotional deprivation. This disorder is reversed by placing the child in a supportive psychosocial milieu. A more detailed description of GH deficiency and growth failure is provided in Chapter 6.

2. **ACTH deficiency**—Monotropic ACTH deficiency is rare and is manifested by the signs and symptoms of adrenocortical insufficiency. LPH deficiency has also been noted in such patients. The defect in these patients may be due to primary failure of the corticotrophs to release ACTH and its related peptide hormones or may be secondary to impaired secretion of CRH by the hypothalamus. Mutations in *Tpit* (T-box 19, a T-box factor only presents in POMC-expressing pituitary lineages) have been described in several patients with early onset isolated ACTH deficiency. Most acquired cases of monotropic ACTH deficiency are now due to checkpoint inhibitor immunotherapy.

3. **Gonadotropin deficiency**—Isolated deficiency of gonadotropins is not uncommon. Kallmann syndrome, initially described in the 1940s, is characterized by an isolated defect in GnRH secretion associated with maldevelopment of the olfactory center with hyposmia or anosmia; X-linked recessive, autosomal dominant, and autosomal recessive patterns of inheritance are seen. Sporadic cases occur, and other neurologic defects such as color blindness and nerve deafness have been reported. At least five Kallmann syndrome genes have been identified: *KAL1, FGFR1, FGF8, PROKR2*, and *PROK2. KAL1* mutations are responsible for the X-linked form of the disease and result in decreased expression of the extracellular glycoprotein anosmin-1. This, in turn, interferes with the normal embryonic development and migration of GnRH-secreting neurons. Because anterior pituitary function is otherwise intact, young men with isolated hypogonadotropic hypogonadism develop a eunuchoid appearance since testosterone deficiency results in failure of epiphysial closure (see Chapter 12). In women, a state of hypogonadotropic hypogonadism manifested by oligomenorrhea or amenorrhea often accompanies weight loss, emotional or physical stress, and athletic training. Anorexia nervosa and marked obesity both result in hypothalamic dysfunction and impaired gonadotropin secretion. Hypothalamic hypogonadism has also been observed in overtrained male athletes. Sickle cell anemia also causes hypogonadotropic hypogonadism due to hypothalamic dysfunction and results in delayed puberty. Clomiphene treatment has been effective in some cases. Isolated gonadotropin deficiency may also be seen in the polyglandular autoimmune syndrome; this deficiency is related to selective pituitary gonadotrope failure from autoimmune hypophysitis. Other chronic illnesses (eg, poorly controlled diabetes, malnutrition) may result in gonadotropin deficiency. Isolated deficiencies of both LH and FSH without an obvious cause such as those described have been reported but are rare. In addition, acquired partial gonadotropin deficiency may occur in middle-aged men. The cause and exact frequency of this disorder are unknown.

4. **TSH deficiency**—Monotropic TSH deficiency is rare and can be caused by a reduction in either hypothalamic TRH secretion (tertiary hypothyroidism) or pituitary TSH secretion (secondary hypothyroidism). These defects have reported in association with gene mutations, empty sella, lymphocytic hypophysitis, and pituitary tumors. Some patients with chronic renal failure also appear to have impaired TSH secretion.

5. **Prolactin deficiency**—PRL deficiency almost always indicates severe intrinsic pituitary damage, and panhypopituitarism is usually present. However, isolated PRL deficiency has been reported after lymphocytic hypophysitis. Deficiencies of TSH and PRL have been noted in patients with pseudohypoparathyroidism.

6. **Multiple hormone deficiencies isolated from other pituitary damage**—Multiple hormone deficiencies result from abnormal pituitary development related to abnormalities of the genes encoding the transcription factors, Pit-1 (TSH, GH, and PRL) and Prop-1 (TSH, GH, PRL, LH, FSH, and ACTH). Table 4–10 lists a number of other genes associated with multiple pituitary hormone deficiencies and hypoplasia including *NR0B1, Rpx, Lhx3, Lhx4*, and *Pitx2*.

Clinical Features

The onset of pituitary insufficiency is usually gradual, and the classic course of progressive hypopituitarism is an initial loss of GH and gonadotropin secretion followed by deficiencies of TSH, then ACTH, and finally PRL.

A. Symptoms   Impairment of GH secretion causes decreased growth in children but may be clinically occult in adult patients. GH deficiency is associated with a decreased sense of well-being and a lower health-related quality of life. Decreased muscle mass and increased fat mass can also be seen, although this may be difficult to discern in any given individual.

Hypogonadism, manifested by amenorrhea in women and decreased libido or erectile dysfunction in men, may antedate the clinical appearance of a hypothalamic-pituitary lesion. The only symptom of PRL deficiency is failure of postpartum lactation.

Hypothyroidism caused by TSH deficiency generally simulates the clinical changes observed in primary thyroid failure; however, it is usually less severe, and goiter is absent. Cold intolerance, dry skin, mental dullness, bradycardia, constipation, hoarseness, and anemia have all been observed; gross myxedematous changes are uncommon.

ACTH deficiency causes adrenocortical insufficiency, and its clinical features resemble those of primary adrenal failure. Weakness, nausea, vomiting, anorexia, weight loss, fever, and hypotension may occur. The zona glomerulosa and the renin-angiotensin system are usually intact; therefore, the dehydration and sodium depletion seen in Addison disease are uncommon. However, these patients are susceptible to hypotension, shock, and cardiovascular collapse, because glucocorticoids are necessary to maintain vascular reactivity, especially during stress. In addition, because of their gradual onset, the symptoms of secondary adrenal insufficiency may go undetected for prolonged periods, becoming manifest only during periods of stress. Hypoglycemia aggravated by GH deficiency may occur with fasting and has been the initial presenting feature of some patients with isolated ACTH deficiency. In contrast to the hyperpigmentation that occurs during states of ACTH excess (Addison disease, Nelson syndrome), depigmentation and diminished tanning have been described as a result of

TABLE 4–10 Genetics of pituitary tumors and pituitary insufficiency.

Pituitary adenomas			
Name	**Gene**	**Tumors/Hormone Excess**	**Function**
MEN1 syndrome	MEN1	Prolactinomas (ACTH, GH, non-functioning tumors less common)	Menin (tumor suppressor)
MEN4 syndrome	CDKN1B	Similar to MEN1	Cyclin-dependent kinase inhibitor 1B (p27, Kip1); important regulator of cell cycle progression, involved in G1 arrest and other functions
Familial isolated pituitary adenomas (familial acromegaly/gigantism)	AIP	GH secreting tumors	Aryl hydrocarbon receptor interacting protein (tumor suppressor)
McCune-Albright syndrome	GNAS1	GH hypersecretion	Adenylyl cyclase
Carney complex	PRKAR1A	GH hypersecretion	Loss of protein kinase A regulatory subunit-1 alpha activity (tumor suppressor)
X-linked acrogigantism (X-LAG)	GPR101	GH secreting tumors	Genomic duplication of Xq26.3
Cushing's disease	USP8	ACTH secreting tumors	Ubiquitin-specific protease 8 (deubiquitinase activity)
DICER1 syndrome	DICER-1	Pituitary blastoma	DICER-1 protein (RNA processing endoribonuclease)
Familial hyperprolactinemia	PRLR	Prolactin hypersecretion	Loss of function mutation in the prolactin receptor
Pituitary adenomas	PTTG1	All pituitary tumor subtypes	Pituitary tumor-transforming gene-1 (securin which regulates sister chromatid separation)
Other	CCND1 (Cyclin D1), H-ras, c-myc, Ptx	Pituitary adenomas/carcinomas	Activators of growth and cell cycle progression in pituitary cells.
Craniopharyngioma	CTNNB1	Adamantinomatous tumors	β-catenin protein, regulator of canonical Wnt signaling pathway
Craniopharyngioma	BRAF	Papillary tumors	B-Raf protein
Pituitary insufficiency			
Name	**Gene**	**Hormone Deficiency**	**Function**
X-linked congenital adrenal hypoplasia	NR0B1	Pituitary and hypothalamus (in addition to adrenals)	DAX-1 nuclear receptor protein (dosage-sensitive sex reversal, adrenal hypoplasia gene on the X chromosome, gene 1)
Early onset isolated ACTH deficiency	Tpit	ACTH deficiency	Transcription factor required for expression of the POMC gene
Kallmann syndrome	KAL1, FGFR1, FGF8, PROKR2, PROK2	GnRH deficiency (low FSH and LH)	Anosmin-1, fibroblast growth factor receptor 1, fibroblast growth factor-8, prokineticin receptor-2, prokineticin-2
Combined pituitary hormone insufficiency	Prop1	Multiple pituitary hormone deficiencies	Prop1 (Prophet of Pit1) protein is a transcription factor
Congenital hypopituitarism	POU1F1 (Pit1)	Multiple pituitary hormone deficiencies	Anterior pituitary transcription factor
Other	Rpx, Lhx3, Lhx4, Pitx2	Pituitary hypoplasia with pituitary hormone deficits	

ACTH insufficiency. In addition, lack of ACTH-stimulated adrenal androgen secretion will cause a decrease in body hair if gonadotropin deficiency is also present.

B. Signs   Abnormal findings on physical examination may be subtle and require careful observation. Patients with hypopituitarism are not cachectic. A photograph of a cachectic patient with "Simmonds syndrome" that appeared in some older textbooks of endocrinology caused confusion. That particular patient probably suffered from anorexia nervosa and was found to have a normal pituitary gland at postmortem examination.

Patients with hypopituitarism may have fine, pale, smooth skin and fine wrinkling of the face. Body and pubic hair may be deficient or absent, and atrophy of the genitalia may occur. Postural hypotension, bradycardia, decreased muscle strength, and delayed deep tendon reflexes occur in more severe cases. Neuro-ophthalmologic abnormalities depend on the presence of a large intrasellar or parasellar lesion.

C. Laboratory and other findings   These may include anemia (related to thyroid and androgen deficiency and chronic disease), hypoglycemia, hyponatremia (related to hypothyroidism and hypoadrenalism, which cause inappropriate water retention, not sodium loss), hypercalcemia (related to adrenal insufficiency with volume contraction and hemoconcentration), and low-voltage bradycardia on electrocardiographic testing. Hyperkalemia, which is common in primary adrenal failure, is

not present. Adult GH deficiency is associated with decreased red blood cell mass, increased LDL cholesterol, and decreased bone mass.

Diagnosis

A. Assessment of target gland function (Figure 4–18)   If endocrine hypofunction is suspected, pituitary hormone

Clinical history and examination

Endocrine-hypofunction suspected[a]

Thyroid function tests
Serum testosterone[b]
IGF-1

Rapid ACTH
stimulation test

Normal

Low

No disease

Impaired cortisol response

Serum TSH, LH, FSH

Plasma ACTH

Elevated

Primary target
gland failure

Low or normal

Low or normal

Hypopituitarism

Insulin hypoglycemia, TRH, GnRH,
or clomiphene tests

Confirm diagnosis, extent of hypofunction,
and level of abnormality

[a]Evaluate GH in children (see text).
[b]Measure prolactin if hypogonadism suspected.

FIGURE 4–18 Diagnostic evaluation of hypothalamic-pituitary-target gland hypofunction.

deficiencies must be distinguished from primary failure of the thyroid, adrenals, or gonads. Basal determinations of each anterior pituitary hormone are useful only if compared to target gland secretion. Baseline laboratory studies should include thyroid function tests (free T_4) and determination of serum testosterone levels. Testosterone is a sensitive indicator of hypopituitarism in women as well as in men. In women, a substantial decrease in testosterone is commonly observed in pituitary failure related to hypofunction of the two endocrine glands responsible for its production—the ovary and the adrenal. Adrenocortical reserve should initially be evaluated by a rapid ACTH stimulation test.

B. Evaluation of prolactin   Because hyperprolactinemia (discussed later), regardless of its cause, leads to gonadal dysfunction, serum PRL should be measured early in the evaluation of hypogonadism.

C. Differentiation of primary and secondary hypofunction
Subnormal thyroid function as shown by appropriate tests, a low serum testosterone level, or an impaired cortisol response to the rapid ACTH stimulation test requires measurement of basal levels of specific pituitary hormones. In primary target gland hypofunction, such as autoimmune polyglandular syndromes types 1 and 2 (APS 1 and 2), TSH, LH, FSH, or ACTH will be elevated. Low or normal values for these pituitary hormones suggest hypothalamic-pituitary dysfunction.

D. Stimulation tests   Provocative endocrine testing may then be employed to confirm the diagnosis and to assess the extent of hypofunction. At present, these tests are not required in most patients.

Treatment

A. ACTH   Treatment of secondary adrenal insufficiency, like that of primary adrenal failure, must include glucocorticoid support (see Chapter 9). Hydrocortisone (15-25 mg/d orally) or prednisone (5-7.5 mg/d orally) in two or three divided doses provides adequate glucocorticoid replacement for most patients. The minimum effective dosage should be given in order to avoid iatrogenic hypercortisolism. Increased dosage is required during periods of stress such as illness, surgery, or trauma. Patients with only partial ACTH deficiency may need steroid treatment only during stress. A two- to threefold increase in steroid dosage during the stressful situation should be recommended, followed by gradual tapering as the stress resolves. Unlike primary adrenal insufficiency, ACTH deficiency does not usually require mineralocorticoid therapy. Patients with adrenal insufficiency should wear medical alert bracelets so they may receive prompt treatment in case of emergency.

B. TSH   The management of patients with secondary hypothyroidism must be based on clinical grounds and the circulating concentration of serum thyroxine (see Chapter 7). The treatment of secondary and tertiary hypothyroidism is identical to that for primary thyroid failure. Full oral replacement dose of levothyroxine sodium is 1.6 µg/kg daily (0.1-0.15 mg/d is usually adequate). Response to therapy is monitored clinically and with measurement of serum-free thyroxine levels, which should be maintained in the mid to upper range of normal. Measurement of TSH levels is obviously of no value in the management of these patients.

Caution: Because thyroid hormone replacement in patients with hypopituitarism may aggravate even partial adrenal insufficiency, the adrenal hormone deficiency should be treated first.

C. Gonadotropins   The object of treatment of secondary hypogonadism is to replace sex steroids and restore fertility (see Chapters 12 and 13).

1. **Estrogens and progesterone**—In premenopausal women, estrogen replacement is essential. Adequate estrogen treatment maintains secondary sex characteristics (eg, vulvar and vaginal lubrication), prevents osteoporosis, and abolishes vasomotor symptoms, with an improvement in sense of well-being. Many estrogen preparations are available (eg, oral estradiol, 1-2 mg daily; conjugated estrogens, 0.3-1.25 mg orally daily; or transdermal estradiol, 0.05-0.1 mg daily). Estrogens should be cycled with a progestin compound (eg, medroxyprogesterone, 5-10 mg orally daily during the last 10 days of estrogen therapy each month) to induce withdrawal bleeding and prevent endometrial hyperplasia. Many oral contraceptive pill combinations are also clinically available.

2. **Ovulation induction**—Ovulation can often be restored in women with hypothalamic-pituitary dysfunction (see Chapter 13). In patients with gonadal failure of hypothalamic origin, clomiphene citrate may cause a surge of gonadotropin secretion resulting in ovulation. Pulsatile subcutaneous injections of GnRH with an infusion pump can also be used to induce ovulation and fertility in women with hypothalamic dysfunction. Combined treatment with FSH (human menopausal gonadotropins; menotropins) and LH (chorionic gonadotropin) can be utilized to provoke ovulation in women with intrinsic pituitary failure. This form of therapy is expensive, and multiple births are a risk (see Chapter 13).

3. **Androgens in women**—Because of a deficiency of both ovarian and adrenal androgens, some women with hypopituitarism have diminished libido despite adequate estrogen therapy. Although experience is limited, small doses of long-acting androgens (testosterone enanthate, 25-50 mg intramuscularly every 4-8 weeks) may be helpful in restoring sexual activity without causing hirsutism. In addition, some reports have suggested that oral dehydroepiandrosterone (DHEA) in doses of 25 to 50 mg/d may restore plasma testosterone levels to normal. A transdermal delivery system is being evaluated for use in women, but efficacy appears to be modest and the long-term safety is unknown.

4. **Androgens in men**—The treatment of male hypogonadism is discussed in Chapter 12. Testosterone gels (administered in doses of 10 to 120 mg daily depending on the concentration) are available in packets and metered-dose pumps. Testosterone patches (available as 2 and 4 mg) can be used in doses up to 6 mg daily. Other therapeutic preparations include intramuscular testosterone enanthate or cypionate in doses of 50 to 100 mg every week or 100 to 200 mg every 2 weeks.

Testosterone undecanoate is an intramuscular preparation (250 mg/mL) available in several countries that can be given every 3 months. Oral testosterone preparations available in the United States are rarely used out of concern for hepatic side effects.

5. **Spermatogenesis**—Spermatogenesis can be achieved in many patients with the combined use of hCG and recombinant FSH. If pituitary insufficiency is of recent onset, therapy with hCG alone may restore both fertility and adequate gonadal steroid production. Pulsatile GnRH infusion pumps have also been used to restore fertility in male patients with secondary hypogonadism.

D. Growth hormone (also see Chapter 6)   Human GH (hGH) produced by recombinant DNA technology is available for use in children with hypopituitarism and for adults with GH deficiency and known pituitary disease. Some studies indicate improvement in body composition, bone mineral density, psychologic well-being, and functional status with GH therapy. However, the long-term benefits and risks remain to be established. In adults, GH is usually administered subcutaneously, once per day in a dosage of 2 to 5 µg/kg. Monitoring of effectiveness is accomplished by measurement of IGF-1, and the dosage of GH is adjusted accordingly (up to about 10 µg/kg/d). Side effects (eg, edema, paresthesias, arrhythmias, glucose intolerance, diabetes) should be assessed. Contraindications to GH therapy include the presence of diabetic retinopathy, active malignancy, intracranial hypertension, radiotherapy, or airway obstruction in individuals with Prader-Willi syndrome.

PITUITARY ADENOMAS

Advances in endocrinologic and neuroradiologic research in recent years have allowed earlier recognition and more successful therapy of pituitary adenomas. In a population-based study, the distribution of each pituitary adenoma subtype was 57% for prolactinomas, 28% for nonfunctioning tumors, 11% for GH-secreting tumors, 2% for corticotroph adenomas, and 2% for unknown functional status. Hypersecretion of TSH, the gonadotropins, or alpha subunits is unusual, although the latter two may account for some of the "nonfunctional" tumors. The differential diagnosis of nonpituitary sellar and parasellar masses is listed in Table 4–11.

Early clinical recognition of the endocrine effects of excessive pituitary secretion, especially the observation that PRL excess causes secondary hypogonadism, has led to early diagnosis of pituitary tumors before the appearance of late manifestations, such as sellar enlargement, panhypopituitarism, and suprasellar extension with visual impairment.

Pituitary **microadenomas** are defined as intrasellar adenomas less than 1 cm in diameter that present with manifestations of hormonal excess without sellar enlargement or extrasellar extension. Panhypopituitarism does not occur, and such tumors can usually be treated successfully.

Pituitary **macroadenomas** are those larger than 1 cm in diameter and cause generalized sellar enlargement. Tumors 1 to 2 cm in diameter confined to the sella turcica can usually be successfully

TABLE 4–11 Differential diagnosis of sellar and parasellar masses (excluding pituitary adenomas).

Benign tumors
Craniopharyngioma
Rathke cleft cyst
Meningioma
Enchondroma
Arachnoid cyst
Dermoid cyst
Empty sella
Nonadenomatous pituitary hyperplasia
Lactotroph hyperplasia during pregnancy
Somatotroph hyperplasia from ectopic GHRH-secreting tumor
Thyrotroph, gonadotroph hyperplasia
Malignant tumors
Sarcoma
Chordoma
Germ cell tumor (ectopic pinealoma)
Metastatic lesions (breast, lung, and other tumors)
Gliomas
Optic glioma
Astrocytoma
Oligodendroglioma
Ependymoma
Vascular lesions
Granulomatous, inflammatory, infectious
Lymphocytic hypophysitis
Sarcoidosis
Histiocytosis X
Tuberculosis
Pituitary abscess

treated; however, larger tumors—especially those with suprasellar, sphenoid sinus, or lateral extensions—are much more difficult to manage. Panhypopituitarism and visual loss increase in frequency with tumor size and suprasellar extension.

Insights into the pathogenesis and biologic behavior of pituitary tumors have been gained from studies of pituitary tumor clonality and somatic mutations. Analyses of allelic X inactivation of specific genes has shown that most pituitary adenomas are monoclonal, a finding most consistent with a somatic mutation model of tumorigenesis; polyclonality of tumors would be expected if tonic stimulation by hypothalamic-releasing factors were the mechanism underlying neoplastic transformation. In fact, transgenic animals expressing GHRH have exhibited pituitary hyperplasia but not pituitary adenomas. One somatic mutation has been found in 30% to 40% of GH-secreting tumors (but not in leukocytes from the same patients). Point mutations in the alpha subunit of the GTP-binding protein responsible for activation of adenylyl cyclase result in constitutive stimulation of pituitary cell growth and function. A host of other candidate genes have been described, including the pituitary tumor transforming gene (*PTTG1*). Overexpression of this gene has been found in all pituitary tumor types and may promote tumorigenesis through cell cycle disruption, stimulation of fibroblast growth factor secretion, and abnormal chromatid separation. Several genetic

syndromes are associated with pituitary tumors or pituitary hormone hypersecretion including multiple endocrine neoplasia (MEN) type 1, McCune-Albright syndrome, Carney complex, familial acromegaly/gigantism, and X-linked acrogigantism. Table 4–10 describes a number of known genetic causes of pituitary adenomas/hormone excess.

Treatment

Pituitary adenomas are treated with surgery, irradiation, or medications. The aims of therapy are to block effects of hormone excess, correct hypersecretion of anterior pituitary hormones, to preserve normal secretion of other anterior pituitary hormones, and to remove or suppress the adenoma itself. These objectives are currently achievable in most patients with pituitary microadenomas; however, in the case of larger tumors, multiple therapies are frequently required and may be less successful.

A. Surgical treatment   The transsphenoidal microsurgical approach to the sella turcica, now done endoscopically in most centers, is the procedure of choice; transfrontal craniotomy is required only in the rare patient with massive suprasellar extension of the adenoma. In the transsphenoidal procedure, the surgeon approaches the pituitary from the nasal cavity through the sphenoid sinus, removes the anterior-inferior sellar floor, and incises the dura. The adenoma is selectively removed; normal pituitary tissue is identified and preserved. Success rates approach 90% in patients with microadenomas. Major complications, including postoperative hemorrhage, cerebrospinal fluid leak, meningitis, and visual impairment, occur in less than 5% of patients and are most frequent in patients with large or massive tumors. Transient diabetes insipidus lasting a few days to 1 to 2 weeks occurs in approximately 15%; permanent diabetes insipidus is rare. A transient form of the syndrome of inappropriate secretion of antidiuretic hormone (SIADH) with symptomatic hyponatremia occurs in 10% of patients within 5 to 14 days of transsphenoidal pituitary microsurgery. These abnormalities of water balance can occur within days of each other making medical management difficult. A *triphasic response* of diabetes insipidus, SIADH, and diabetes insipidus is occasionally encountered and thought to be due to early hypothalamic dysfunction, followed by release of ADH from the degenerating pituitary, and then depletion of ADH stores. Surgical hypopituitarism is rare in patients with microadenomas but approaches 5% to 10% in patients with larger tumors. The perioperative management of such patients should include glucocorticoid administration in stress doses (see Chapter 9) and postoperative assessment of daily weight, fluid balance, and electrolyte status to look for evidence of diabetes insipidus. Mild diabetes insipidus is managed by giving fluids orally; in more severe cases—urine output greater than 5 to 6 L/24 h—ADH therapy in the form of desmopressin, or DDAVP, can be administered intranasally, orally, subcutaneously, or parenterally (see Chapter 5). SIADH is managed by fluid restriction; however, in more severe cases, hypertonic saline may be required. ADH receptor antagonists (tolvaptan is given orally and conivaptan is administered

intravenously) are also now available for use in some patients (see section on SIADH).

B. Radiotherapy   Pituitary irradiation should be reserved for patients who have had incomplete resection of larger pituitary adenomas and whose tumors are not amenable to, or have failed, medical therapy.

1. **Conventional irradiation**—Conventional irradiation using high-energy sources, in total doses of 4000 to 5000 cGy given in daily doses of 180 to 200 cGy, is most commonly employed. The response to radiation therapy is slow, and 5 to 10 years may be required to achieve the full effect (see section on Acromegaly). Treatment is ultimately successful in about 80% of patients with acromegaly but in only about 55% to 60% of patients with Cushing disease. The response rate in prolactinomas is not precisely known, but tumor progression is prevented in most patients. Morbidity during radiotherapy is minimal, although some patients experience malaise and nausea, and serous otitis media may occur. Hypopituitarism is common, and the incidence increases with time following radiotherapy—about 50% to 60% at 5 to 10 years. Rare late complications include damage to the optic nerves and chiasm, seizures, and radionecrosis of brain tissue. Recent studies from the United Kingdom have shown that conventional radiotherapy is a major risk factor for excess mortality in acromegaly.
2. **Gamma knife radiosurgery**—This form of radiotherapy utilizes stereotactic CT-guided cobalt-60 gamma radiation to deliver high-radiation doses to a narrowly focused area. Remission rates have been reported in the range of 43% to 78%. An adequate distance of the pituitary tumor from the optic chiasm is needed to prevent radiation-induced damage. Repeat treatments put patients at higher risk of new visual or third, fourth, or sixth cranial nerve deficits.
3. **Proton stereotactic radiotherapy**—Experience with this modality is limited. In one study, 52% of patients with Cushing disease had clinical remission. A similar rate of new pituitary hormone deficiencies was seen, although no visual complications or brain injury was reported.

C. Medical treatment—   Medical management of pituitary adenomas became feasible with the availability of bromocriptine, a dopamine agonist that suppresses both PRL and tumor growth in patients with prolactinomas. A number of medications have been used to treat hypercortisolism associated with Cushing's disease including ketoconazole, metyrapone, mitotane, etomidate, cabergoline, pasireotide, and mifepristone. Somatostatin analogs are useful in the therapy of acromegaly and some TSH-secreting adenomas. Specifics of the use of these and other medications are discussed later.

Posttreatment Follow-Up

Patients undergoing transsphenoidal microsurgery should be reevaluated 4 to 8 weeks postoperatively to document that complete removal of the adenoma and correction of endocrine hypersecretion has been achieved. Prolactinomas are assessed by basal PRL measurements, GH-secreting tumors by glucose suppression testing and IGF-1 levels, and ACTH-secreting adenomas by

measurement of late-night salivary cortisol, urine-free cortisol, and the response to low-dose dexamethasone suppression (see later). Other anterior pituitary hormones—TSH and LH/FSH—should also be assessed as described earlier in the section on endocrine evaluation. In patients with successful responses, yearly evaluation should be performed to watch for late recurrence; late hypopituitarism does not occur after microsurgery. MRI is not necessary in patients with normal postoperative pituitary function but should be utilized in patients with persistent or recurrent disease.

Follow-up of patients treated by pituitary irradiation is also essential, because the response to therapy may be delayed and the incidence of hypopituitarism increases with time. Yearly endocrinologic assessment of both the hypersecreted hormone and the other pituitary hormones is recommended.

1. PROLACTINOMAS

PRL hypersecretion is the most common endocrine abnormality due to hypothalamic-pituitary disorders, and PRL is the hormone most commonly secreted in excess by pituitary adenomas.

The understanding is that PRL hypersecretion causes not only galactorrhea but also gonadal dysfunction, and the use of PRL measurements in screening such patients has permitted recognition of these PRL-secreting tumors before the development of sellar enlargement, hypopituitarism, or visual impairment. Thus, plasma PRL should be measured in patients with galactorrhea, suspected hypothalamic-pituitary dysfunction, or sellar enlargement and in those with unexplained gonadal dysfunction, including amenorrhea, infertility, decreased libido, or impotence (Table 4–12).

Pathology

PRL-secreting pituitary adenomas arise most commonly from the lateral wings of the anterior pituitary, but with progression they fill the sella turcica and compress the normal anterior and posterior lobes. Tumor size varies greatly from microadenomas to large invasive tumors with extrasellar extension. Most patients have microadenomas (ie, tumors <1 cm in diameter at diagnosis).

Prolactinomas usually appear chromophobic on routine histologic study, reflecting the inadequacy of the techniques used. The cells are small and uniform, with round or oval nuclei and scanty cytoplasm, and secretory granules are usually not visible with routine stains. The stroma contains a diffuse capillary network.

Electron microscopic examination shows that prolactinoma cells characteristically contain secretory granules that usually

TABLE 4–12 Indications for prolactin measurement.

Galactorrhea
Enlarged sella turcica
Suspected pituitary tumor
Hypogonadotropic hypogonadism
Unexplained amenorrhea
Unexplained male hypogonadism or infertility

range from 100 to 500 nm in diameter and are spherical. Larger granules (400-500 nm), which are irregular or crescent-shaped, are less commonly seen. The cells show evidence of secretory activity, with a large Golgi area, nucleolar enlargement, and a prominent endoplasmic reticulum. Immunocytochemical studies of these tumors have confirmed that the secretory granules indeed contain PRL.

Clinical Features

The clinical manifestations of PRL excess are the same regardless of the cause (see later). The classic features are galactorrhea and amenorrhea in women and decreased libido or impotence in men. Although the sex distribution of prolactinomas is approximately equal, microadenomas are much more common in women, presumably because of earlier recognition of the endocrine consequences of PRL excess.

A. Galactorrhea Galactorrhea occurs in the majority of women with prolactinomas and is much less common in men. It is usually not spontaneous, or may be present only transiently or intermittently. The absence of galactorrhea despite markedly elevated PRL levels is probably due to concomitant deficiency of the gonadal hormones required to initiate lactation (see Chapter 16).

B. Gonadal dysfunction

1. **In women**—Amenorrhea, oligomenorrhea with anovulation, or infertility is present in approximately 90% of women with prolactinomas. These menstrual disorders usually present concurrently with galactorrhea, if it is present, but may either precede or follow it. The amenorrhea is usually secondary and may follow pregnancy or oral contraceptive use. Primary amenorrhea occurs in the minority of patients who have onset of hyperprolactinemia during adolescence. The necessity of measuring PRL in patients with unexplained primary or secondary amenorrhea is emphasized by several studies showing that hyperprolactinemia occurs in as many as 20% of patients with neither galactorrhea nor other manifestations of pituitary dysfunction. A number of these patients have been shown to have prolactinomas.

Gonadal dysfunction in these women is due to interference with the hypothalamic-pituitary-gonadal axis by the hyperprolactinemia and, except in patients with large or invasive adenomas, is not due to destruction of the gonadotropin-secreting cells. This has been documented by the return of menstrual function following reduction of PRL levels to normal by drug treatment or surgical removal of the tumor. Although basal gonadotropin levels are frequently within the normal range despite reduction of sex steroid levels in hyperprolactinemic patients, PRL inhibits both the normal pulsatile secretion of LH and FSH and the midcycle LH surge, resulting in anovulation. The positive feedback effect of estrogen on gonadotropin secretion is also inhibited; in fact, patients with hyperprolactinemia are usually estrogen-deficient.

Estrogen deficiency in women with prolactinomas may be accompanied by decreased vaginal lubrication, other symptoms of estrogen deficiency, and low bone mass as assessed by bone densitometry. Other symptoms may include weight gain, fluid retention, and irritability. Hirsutism may also occur,

accompanied by elevated plasma levels of DHEA sulfate. Patients with hyperprolactinemia may also suffer from anxiety and depression. In small clinical trials, treatment with dopamine agonists has been shown to improve psychologic well-being in such patients.

2. **In men**—In men, PRL excess may also occasionally cause galactorrhea; however, the usual manifestations are those of hypogonadism. The initial symptom is decreased libido, which may be dismissed by both the patient and physician as due to psychosocial factors; thus, the recognition of prolactinomas in men is frequently delayed, and marked hyperprolactinemia (PRL >200 ng/mL [9.1 nmol/L]) and sellar enlargement are usual. Unfortunately, prolactinomas in men are often not diagnosed until late manifestations such as headache, visual impairment, or hypopituitarism appear; virtually all such patients have a history of sexual or gonadal dysfunction. Serum testosterone levels are low, and in the presence of normal or subnormal gonadotropin levels, PRL excess should be suspected as should other causes of hypothalamic-pituitary-gonadal dysfunction (see section on Hypopituitarism). Impotence also occurs in hyperprolactinemic males. Its cause is unclear, because testosterone replacement may not reverse it if hyperprolactinemia is not corrected. Male infertility accompanied by reduction in sperm count is a less common initial complaint.

C. Tumor progression   In general, the growth of prolactinomas is slow; several studies have shown that most microadenomas do not change significantly in size, and macroadenomas tend to grow very slowly.

Differential Diagnosis

The many conditions associated with hyperprolactinemia are listed in Table 4–8. Pregnancy, hypothalamic-pituitary disorders, primary hypothyroidism, and drug ingestion are the most common causes.

Hypothalamic lesions frequently cause PRL hypersecretion by decreasing the secretion of dopamine that tonically inhibits PRL release; the lesions may be accompanied by panhypopituitarism. Similarly, traumatic or surgical section of the pituitary stalk leads to hyperprolactinemia and hypopituitarism. Nonfunctional pituitary macroadenomas frequently cause mild hyperprolactinemia by compression of the pituitary stalk or hypothalamus.

Pregnancy leads to a physiologic increase in PRL secretion; the levels increase as pregnancy continues and may reach 200 ng/mL (9.1 nmol/L) during the third trimester. Following delivery, basal PRL levels gradually fall to normal over several weeks but increase in response to breast feeding. Hyperprolactinemia persisting for 6 to 12 months or longer following delivery is an indication for evaluation. PRL levels are also high in normal neonates.

Several systemic disorders lead to hyperprolactinemia. Primary hypothyroidism is a common cause, and measurement of thyroid function, and especially TSH, should be part of the evaluation. In primary hypothyroidism, there is hyperplasia of both thyrotrophs and lactotrophs, presumably due to TRH hypersecretion. This may result in significant pituitary gland enlargement, which may be mistaken for a PRL-secreting pituitary tumor. The PRL response to TRH is usually exaggerated in these patients. PRL

may also be increased in liver disease, particularly in patients with severe cirrhosis, and in patients with chronic renal failure.

PRL excess and galactorrhea may also be caused by breast disease, nipple stimulation, disease or injury to the chest wall, and spinal cord lesions. These disorders increase PRL secretion by stimulation of afferent neural pathways.

Artifactual elevations in PRL levels may be observed in the presence of anti-PRL antibodies or of macroprolactinemia. In the latter, a high-molecular-weight complex of PRL molecules maintains immunologic activity, but minimal or no bioactivity. Macroprolactinemia can be assessed with polyethylene glycol precipitation of serum samples.

The most common cause of hyperprolactinemia is drug ingestion, and a careful history of drug intake must be obtained. Elevated PRL levels, galactorrhea, and amenorrhea may occur following estrogen therapy or oral contraceptive use, but their persistence should suggest prolactinoma. Many other drugs also cause increased PRL secretion and elevated plasma levels (see Table 4–8). PRL levels are usually less than 200 ng/mL (9 nmol/L), and the evaluation should focus on discontinuation of the drug or medication and reevaluation of the patient after several weeks. In patients in whom drug withdrawal is not feasible, neuroradiologic studies, if normal, usually exclude prolactinoma.

Diagnosis

A. General evaluation   The evaluation of patients with galactorrhea or unexplained gonadal dysfunction with normal or low plasma gonadotropin levels should first include a history regarding menstrual status, pregnancy, fertility, sexual function, and symptoms of hypothyroidism or hypopituitarism. Current or previous use of medication, drugs, or estrogen therapy should be documented. Basal PRL levels, gonadotropins, thyroid function tests, and TSH levels should be established, as well as serum testosterone in men. Liver and kidney function should be assessed. A pregnancy test should be performed in women with amenorrhea.

Patients with galactorrhea but normal menses may not have hyperprolactinemia and usually do not have prolactinomas. If the PRL level is normal, they may be reassured and followed with sequential PRL measurements. Those with elevated levels require further evaluation as described later.

B. Specific diagnosis   When other causes of hyperprolactinemia have been excluded, the most likely cause of persistent hyperprolactinemia is a prolactinoma, especially if there is associated hypogonadism. Because currently available suppression and stimulation tests do not distinguish PRL-secreting tumors from other causes of hyperprolactinemia, the diagnosis must be established by the assessment of both basal PRL levels and neuroradiologic studies. Patients with large tumors and marked hyperprolactinemia usually present little difficulty. With very rare exceptions, basal PRL levels greater than 200 to 300 ng/mL (9.1-13.7 nmol/L) are virtually diagnostic of prolactinoma. In addition, because there is a general correlation between the PRL elevation and the size of the pituitary adenoma, these patients

usually have sellar enlargement and obvious macroadenomas. Similarly, if the basal PRL level is between 100 and 200 ng/mL (4.5 and 9.1 nmol/L), the cause is usually prolactinoma. These patients may have either micro- or macroadenomas; however, with basal levels of PRL greater than 100 ng/mL (4.5 nmol/L), the PRL-secreting tumor is usually radiologically evident, and again the diagnosis is generally straightforward. Patients with mild to moderate hyperprolactinemia (20-100 ng/mL [0.9-4.5 nmol/L]) present the greatest difficulty in diagnosis, because both PRL-secreting microadenomas and the many other conditions causing hyperprolactinemia (see Table 4–8) cause PRL hypersecretion of this degree. In such patients, MRI frequently demonstrates a definite pituitary microadenoma. Scans showing only minor or equivocal abnormalities should be interpreted with caution, because of the high incidence of false-positive scans in the normal population (see Neuroradiologic Evaluation, earlier). Because the diagnosis cannot be either established or excluded in patients with normal or equivocal neuroradiologic studies, they require further evaluation or serial assessment (see later). Dilutions of prolactin samples can be performed in patients with modest prolactin elevation or macroadenomas to rule out interference from the high-dose *hook effect*; large quantities of antigen can impair antigen-antibody binding, resulting in erroneously low prolactin measurements in some immunoassays.

Treatment

Satisfactory control of PRL hypersecretion, cessation of galactorrhea, and return of normal gonadal function can be achieved in most patients with PRL-secreting microadenomas. In patients with hyperprolactinemia, ovulation should not be induced without careful assessment of pituitary anatomy, because pregnancy may cause further expansion of these tumors as discussed later.

Although most microadenomas do not progress, treatment of these patients is recommended to restore normal estrogen levels and fertility and to prevent early osteoporosis secondary to persistent hypogonadism. In addition, medical or surgical therapy is more successful in these patients than in those with larger tumors. All patients with PRL-secreting macroadenomas should be treated, because of the risks of further tumor expansion, hypopituitarism, and visual impairment. Patients with persistent hyperprolactinemia and hypogonadism and normal neuroradiologic studies—that is, those in whom prolactinoma cannot be definitely established—may be managed by observation if hypogonadism is of short duration. However, in patients whose hypogonadism has persisted for more than 6 to 12 months, dopamine agonists should be used to suppress PRL secretion and restore normal gonadal function. In women with macroprolactinomas, replacement estrogen therapy should be initiated only after PRL hypersecretion has been controlled by dopamine agonist therapy since estrogen stimulates lactotroph hyperplasia and may increase tumor size. In this regard, periodic measurement of serum PRL should be performed in women with microadenomas on estrogen therapy who are not also receiving dopamine agonist therapy.

A. Dopamine agonists   Bromocriptine became available in the United States more than 30 years ago and was the first effective medical therapy for pituitary adenomas; however, cabergoline is more potent, much longer acting, and better tolerated. Cabergoline has, therefore, become the dopamine agonist of choice in the therapy of prolactinomas.

1. **Bromocriptine**—Bromocriptine stimulates dopamine receptors and has effects at both the hypothalamic and pituitary levels. It is effective therapy for a PRL-secreting pituitary adenoma and directly inhibits PRL secretion by the tumor. Doses of 2.5 to 5 mg daily are often effective.

2. **Cabergoline**—Cabergoline, a newer nonergot dopamine agonist, is administered once or twice a week and has a better side-effect profile than bromocriptine. It is as effective as bromocriptine in reducing macroadenoma size and is more effective in reducing PRL levels. It has been used successfully in most patients previously intolerant or resistant to bromocriptine. Cabergoline should be started at a dosage of 0.25 to 0.50 mg twice per week. Bedtime administration with a snack seems to reduce side effects. Although several studies demonstrated associations between cabergoline and cardiac valve disease in patients treated for Parkinson disease, clinically relevant valve disease in patients treated for prolactinoma appears to be rare, likely because doses used to manage symptoms of Parkinsonism are often 20 to 30 times higher and administered daily rather than twice weekly. The risk also appears to be related to the affinity of different dopamine agonists for valvular serotonin (5-HT2B) receptors. Until larger, prospective safety studies are available, some authorities recommend echocardiographic evaluation in patients who are expected to need long-term treatment, especially at high doses.

    a. **Microadenomas**—Cabergoline is successful in about 90% of patients, and very few are intolerant or resistant. Correction of hyperprolactinemia allows recovery of normal gonadal function; ovulation and fertility are restored, so that barrier and non-hormonal forms of contraception should be advised if pregnancy is not desired. Both bromocriptine and cabergoline are now considered to be safe for the restoration of the hypothalamic-pituitary-gonadal axis and ovulation induction. In patients with microadenomas, the risk of major expansion of adenoma during the pregnancy is less than 2%; however, both the patient and the physician must be aware of this potential complication. Current data do not indicate an increased risk of multiple pregnancy, abortion, or fetal malformations in pregnancies occurring in women taking dopamine agonists; however, patients should be instructed to discontinue these drugs at the first missed menstrual period and obtain a pregnancy test. Although no late toxicity has yet been reported other than the side effects noted earlier, questions about possible long-term risk are currently unanswered. Maternal hyperprolactinemia should not be harmful to the developing fetus; prolactin increases normally during pregnancy and does not appear to cross the placenta.

    b. **Macroadenomas**—Dopamine agonists are effective in controlling hyperprolactinemia in patients with PRL-secreting macroadenomas even when basal PRL levels are markedly elevated. Dopamine agonists may be used either as initial therapy or to control residual hyperprolactinemia in patients unsuccessfully treated with surgery or radiotherapy. Dopamine agonists should not be used to induce ovulation and

pregnancy in women with untreated macroadenomas, because the risk of tumor expansion and visual deficits in the later part of pregnancy is approximately 15% to 25%. These patients should first have tumor volume decreased with medical therapy or be treated with surgery prior to induction of ovulation.

Dopamine agonists normalize PRL secretion in about 60% to 70% of patients with macroadenomas and also reduce tumor size in about the same percentage of patients. Reduction of tumor size may occur within days to weeks following institution of therapy. The drugs have been used to restore vision in patients with major suprasellar extension and chiasmal compression.

c. **Long-term remission**—Current studies suggest that 30% to 40% of patients with micro- and macroadenomas will remain in long-term remission after withdrawal of cabergoline therapy of 2 to 3 years duration provided that they have normalization of PRL levels and tumor shrinkage. Larger macroadenomas (>2.0 cm) are likely to recur, and long-term dopamine agonist therapy should be continued in these patients.

B. Surgical treatment   Transsphenoidal microsurgery is the surgical procedure of choice in patients with prolactinomas.

1. **Microadenomas**—In patients with microadenomas, remission, as measured by restitution of normal PRL levels, normal menses, and cessation of galactorrhea, is achieved in 85% to 90% of cases. Success is most likely in patients with basal PRL levels under 200 ng/mL (9.1 nmol/L) and duration of amenorrhea of less than 5 years. In these patients, the incidence of surgical complications is less than 2%, and hypopituitarism is a rare complication. Thus, in this group of patients with PRL-secreting microadenomas, PRL hypersecretion can be corrected, gonadal function restored, and secretion of TSH and ACTH preserved. Recurrence rates vary considerably in reported series. In our experience, approximately 85% of patients have had long-term remissions, and 15% have had recurrent hyperprolactinemia.

2. **Macroadenomas**—Transsphenoidal microsurgery is considerably less successful in restoring normal PRL secretion in patients with macroadenomas; many clinicians would treat these patients with dopamine agonists alone. The surgical outcome is directly related to tumor size and the basal PRL level. Thus, in patients with tumors 1 to 2 cm in diameter without extrasellar extension and with basal PRL levels less than 200 ng/mL (9.1 nmol/L), transsphenoidal surgery is successful in about 80% of cases. In patients with higher basal PRL levels and larger or invasive tumors, the success rate—defined as complete tumor resection and restoration of normal basal PRL secretion—is 25% to 50%. Although progressive visual loss or pituitary apoplexy is a clear indication for surgery, the great majority of these patients should be treated with dopamine agonists.

C. Radiotherapy   Conventional radiation therapy is reserved for patients with PRL-secreting macroadenomas who have persistent hyperprolactinemia and who have not responded to attempts to control their pituitary adenomas with surgery or dopamine agonists. In this group of patients, radiotherapy with 4000 to 5000 cGy prevents further tumor expansion, although

PRL levels usually do not fall into the normal range. Impairment of anterior pituitary function occurs in approximately 50% to 60% of patients. Experience with gamma knife radiosurgery in prolactinomas is limited and rates of remission and reduction in tumor volume have been reported with varying degrees of success.

Selection of Therapy for Prolactinomas

The selection of therapy for prolactinomas depends on the wishes of the patient, the patient's plans for pregnancy and tolerance of medical therapy, and the availability of a skilled neurosurgeon.

A. Microadenomas   Most men and premenopausal women should be treated to prevent the occasional tumor progression, loss of bone mass, and the other effects of prolonged hypogonadism. Medical therapy with cabergoline effectively restores normal gonadal function and fertility, and pregnancy carries only a small risk of tumor expansion. Those patients who respond should be treated for 2 to 3 years, and then the drug should be withdrawn to determine if long-term remission will occur. Patients who have recurrence of hyperprolactinemia after cabergoline withdrawal may resume the drug or choose to have surgical excision. In asymptomatic postmenopausal women who decide against using medical therapy, intermittent monitoring of prolactin as a marker of tumor enlargement should be performed. Transsphenoidal adenectomy, either initially or after a trial of dopamine agonist therapy, carries little risk when performed by an experienced neurosurgeon and offers a high probability of long-term remission.

B. Macroadenomas   Primary surgical therapy in these patients usually does not result in long-term remission, so medical therapy is the primary therapy of choice, particularly when the patient's PRL levels are greater than 200 ng/mL (9.1 nmol/L) and the tumor is larger than 2 cm. Although transsphenoidal microsurgery rapidly decreases tumor size and decompresses the pituitary stalk, the optic chiasm, and the cavernous sinuses, there is usually residual tumor and hyperprolactinemia. Thus, these patients require additional therapy with dopamine agonists. Although tumor growth and PRL secretion can be controlled by medical therapy in most patients, therapeutic failure can result from drug intolerance, poor compliance, or resistance. Radiation therapy is reserved for postsurgical patients with residual adenomas who are not controlled with dopamine agonists.

2. ACROMEGALY AND GIGANTISM

GH-secreting pituitary adenomas are second in frequency to prolactinomas for functioning pituitary tumors and cause the classic clinical syndromes of acromegaly and gigantism.

The characteristic clinical manifestations are the consequence of chronic GH hypersecretion, which in turn leads to excessive generation of IGF-1, the mediator of most of the effects of GH (see Chapter 6). Although overgrowth of bone is the classic

feature, GH excess causes a generalized systemic disorder with deleterious effects and an increased mortality rate, although deaths are rarely due to the space-occupying or destructive effects of pituitary adenoma per se.

Acromegaly and gigantism are virtually always secondary to a pituitary adenoma. Ectopic GHRH secretion has been identified as another cause of GH hypersecretion and acromegaly in a few patients with carcinoid or islet cell tumors. Reports of intrapituitary GHRH-secreting gangliocytomas in direct contiguity with GH-secreting somatotroph adenomas and a report of a GHRH-secreting hypothalamic hamartoma in a patient with acromegaly provide a link between ectopic and eutopic GHRH production. Ectopic secretion of GH per se is very rare but has been documented in a few lung tumors.

In adults, GH excess leads to acromegaly, the syndrome characterized by local overgrowth of bone, particularly of the skull and mandible. Linear growth does not occur, because of prior fusion of the epiphyses of long bones. In childhood and adolescence, the onset of chronic GH excess leads to gigantism. Most of these patients have associated hypogonadism, which delays epiphysial closure and the combination of IGF-1 excess and hypogonadism leads to a striking acceleration of linear growth. Most patients with gigantism also have features of acromegaly if GH hypersecretion persists through adolescence and into adulthood.

Pathology

Pituitary adenomas causing acromegaly are usually more than 1 cm in diameter when the diagnosis is established. These tumors arise from the lateral wings of the anterior pituitary; less than 10% are diagnosed as microadenomas.

GH-secreting adenomas are of two histologic types: densely and sparsely granulated. However, there appears to be no difference in the degree of GH secretion or clinical manifestations in these patients. About 15% of GH-secreting tumors also contain lactotrophs, and these tumors thus hypersecrete both GH and PRL.

Etiology and Pathogenesis

In most cases, excessive pituitary GH secretion is a primary pituitary disorder. A somatic mutation in the G_s protein leading to excessive cAMP production has been identified in 40% of GH-secreting adenomas. Familial isolated pituitary adenomas (FIPA) is characterized by genetic defects in the aryl hydrocarbon receptor interacting protein (AIP) gene, which probably acts as a tumor suppressor. A recently described syndrome of pituitary gigantism termed X-linked acrogigantism (X-LAG) appears to be due to genomic duplication of Xq26.3 (which includes the *GPR101* gene) (see Table 4–10). Pituitary adenomas are present in virtually all patients and are usually greater than 1 cm in diameter; hyperplasia alone is rare, and nonadenomatous anterior pituitary tissue does not exhibit somatotroph hyperplasia when examined histologically. In addition, there is a return of normal GH levels and dynamic control of GH secretion following selective removal of the pituitary adenoma.

Pathophysiology

In acromegaly, GH secretion is increased and its dynamic control is abnormal. Secretion remains episodic; however, the number, duration, and amplitude of secretory episodes are increased, and they occur randomly throughout the 24-hour period. The characteristic nocturnal surge is absent, and there are abnormal responses to suppression and stimulation. Thus, glucose suppressibility is lost (see Diagnosis, later), and GH stimulation by hypoglycemia is usually absent. TRH and GnRH may cause GH release, whereas these substances do not normally stimulate GH secretion. Dopamine and dopamine agonists such as bromocriptine and apomorphine, which normally stimulate GH secretion, paradoxically cause GH suppression in about 70% to 80% of patients with acromegaly.

Most of the deleterious effects of chronic GH hypersecretion are caused by its stimulation of excessive amounts of IGF-1 (see Chapter 6), and plasma levels of this protein are increased in acromegaly. The growth-promoting effects of IGF-1 (DNA, RNA, and protein synthesis) lead to the characteristic proliferation of bone, cartilage, and soft tissues and increase in size of other organs to produce the classic clinical manifestations of acromegaly. The insulin resistance and carbohydrate intolerance seen in acromegaly appear to be direct effects of GH and not due to IGF-1 excess.

Clinical Features

The sex incidence of acromegaly is approximately equal; the mean age at diagnosis is approximately 40 years; and the duration of symptoms is usually 5 to 10 years before the diagnosis is established.

Acromegaly is a chronic disabling and disfiguring disorder with increased late morbidity and mortality if untreated. Although spontaneous remissions have been described, the course is slowly progressive in the great majority of cases.

A. Symptoms and signs   Early manifestations (Table 4–13) include soft tissue proliferation, with enlargement of the hands and feet and coarsening of the facial features. This is usually accompanied by increased sweating, heat intolerance, oiliness of the skin, fatigue, and weight gain.

At diagnosis, most patients have classic manifestations, and acral and soft tissue changes are always present. Bone and cartilage changes affect chiefly the face and skull. These changes include thickening of the calvarium; increased size of the frontal sinuses, which leads to prominence of the supraorbital ridges; enlargement of the nose; and downward and forward growth of the mandible, which leads to prognathism and widely spaced teeth. Soft tissue growth also contributes to the facial appearance, with coarsening of the features and facial and infraorbital puffiness. The hands and feet are predominantly affected by soft tissue growth; they are large, thickened, and bulky, with blunt, spade-like fingers (Figure 4–19) and toes. A bulky, sweaty handshake frequently suggests the diagnosis, and there are increases in ring, glove, and shoe sizes. There is generalized thickening of the skin, with increased oiliness and sweating. Acne, sebaceous cysts, and fibromata mollusca (skin tags

TABLE 4–13 Clinical manifestations of acromegaly in 100 patients.

Manifestations of GH excess	
Acral enlargement	100[a]
Soft tissue overgrowth	100
Hyperhidrosis	88
Lethargy or fatigue	87
Weight gain	73
Paresthesias	70
Joint pain	69
Photophobia	46
Papillomas	45
Hypertrichosis	33
Goiter	32
Acanthosis nigricans	29
Hypertension	24
Cardiomegaly	16
Renal calculi	11
Disturbance of other endocrine functions	
Hyperinsulinemia	70
Glucose intolerance	50
Irregular or absent menses	60
Decreased libido or impotence	46
Hypothyroidism	13
Galactorrhea	13
Gynecomastia	8
Hypoadrenalism	4
Local manifestations	
Enlarged sella	90
Headache	65
Visual deficit	20

[a]Percentage of patients in whom these features were present.

Adapted with permission from Friesen SR. *Surgical Endocrinology: Clinical Syndromes.* Philadelphia: Lippincott Williams & Wilkins; 1978.

and papillomas) are common, as is acanthosis nigricans of the axillae and neck and hypertrichosis in women.

These bony and soft tissue changes are accompanied by systemic manifestations, which include hyperhidrosis, heat intolerance, lethargy, fatigue, and increased sleep requirement. Sleep apnea, both obstructive and central, is very common in patients with acromegaly. This has particular importance for the anesthesiologist who must take special precautions to protect the airway during surgery. Moderate weight gain usually occurs. Paresthesias, usually due to carpal tunnel compression, occur in 70%; sensorimotor neuropathies occur uncommonly. Bone and cartilage overgrowth leads to arthralgias and in longstanding cases to degenerative arthritis of the spine, hips, and knees. Photophobia of unknown cause occurs in about half of cases and is most troublesome in bright sunlight and during night driving.

GH excess leads to generalized visceromegaly, clinically evident as thyromegaly and enlargement of the salivary glands. Enlargement of other organs is usually not clinically detectable.

Hypertension occurs in about 25% of patients and cardiomegaly in about 15%. Cardiac enlargement may be secondary to hypertension, atherosclerotic disease, or, rarely, to acromegalic cardiomyopathy. Renal calculi occur in 11% secondary to the hypercalciuria induced by GH excess.

Other endocrine and metabolic abnormalities are common and may be due either to GH excess or to mechanical effects of the pituitary adenoma. Glucose intolerance and hyperinsulinism occur in 50% and 70% of patients, respectively, owing to GH-induced insulin resistance. Overt clinical diabetes occurs in a minority, and diabetic ketoacidosis is rare. Hypogonadism occurs in 60% of female and 46% of male patients and is of multifactorial origin; tumor growth and compression may impair pituitary gonadotropin secretion, and associated hyperprolactinemia (see later) or the PRL-like effect of excessive GH secretion may impair gonadotropin and gonadal function. In men, low total plasma testosterone levels may be due to GH suppression of sex hormone–binding globulin (SHBG) levels; in these cases, plasma-free testosterone levels may be normal, with normal gonadal function. With earlier diagnosis, hypothyroidism and hypoadrenalism due to destruction of the normal anterior pituitary are unusual and are

FIGURE 4–19 Markedly increased soft tissue bulk and blunt fingers in a middle-aged man with acromegaly.

present in only 13% and 4% of patients, respectively. Galactorrhea occurs in about 15% and is usually caused by hyperprolactinemia from a pituitary adenoma with a mixed cell population of somatotrophs and lactotrophs. Gynecomastia of unknown cause occurs in about 10% of men. Although acromegaly may be a component of MEN type 1 syndrome, it is distinctly unusual, and concomitant parathyroid hyperfunction or pancreatic islet cell tumors are rare.

When GH hypersecretion is present for many years, late complications occur, including progressive cosmetic deformity and disabling degenerative arthritis. In addition, the mortality rate is increased; after age 45, the death rate in acromegaly from cardiovascular and cerebrovascular atherosclerosis, respiratory diseases, and colon cancer is two to four times that of the normal population. Death rates tend to be higher in patients with hypertension, cardiovascular disease, or clinical diabetes mellitus.

Space-occupying manifestations of the pituitary adenoma are also common in acromegaly (eg, 65% of patients have headache). Although visual impairment was usually present in older series, it now occurs in only 15% to 20%, because most patients are now diagnosed because of the manifestations of GH excess.

B. Laboratory findings   In addition to elevations in IGF-1 and GH, postprandial plasma glucose may be elevated, and serum insulin is increased in 70% of cases. Elevated serum phosphate (due to increased renal tubular resorption) and hypercalciuria appear to be due to direct effects of GH or IGF-1.

C. Imaging studies   Plain films (Figure 4–20) show sellar enlargement in 90% of cases. Thickening of the calvarium, enlargement of the frontal and maxillary sinuses, and enlargement of the jaw can also be seen. Radiographs of the hand show increased soft tissue bulk, "arrowhead" tufting of the distal phalanges, increased width of intra-articular cartilages, and cystic changes of the carpal bones. Radiographs of the feet show similar changes, and there is increased thickness of the heel pad (normal, <22 mm).

Diagnosis

Acromegaly is usually clinically obvious and can be readily confirmed by assessment of GH secretion; basal fasting GH levels (normal, 1-5 ng/mL [46-232 pmol/L]) are greater than 10 ng/mL (465 pmol/L) in more than 90% of patients and range from 5 ng/mL (232 pmol/L) to greater than 500 ng/mL (23,000 pmol/L), with a mean of approximately 50 ng/mL (2300 pmol/L). However, single measurements are not entirely reliable, because GH secretion is episodic in acromegaly and because other conditions may increase GH secretion (see discussed later).

A. Glucose suppression   Suppression with oral glucose is the simplest and most specific dynamic test for acromegaly. In healthy subjects, oral administration of 100 g of glucose causes a reduction of the GH level to less than 1 ng/mL (47 pmol/L) at 60 minutes. In acromegaly, GH levels may decrease, increase, or show no change; however, they do not decrease to less than
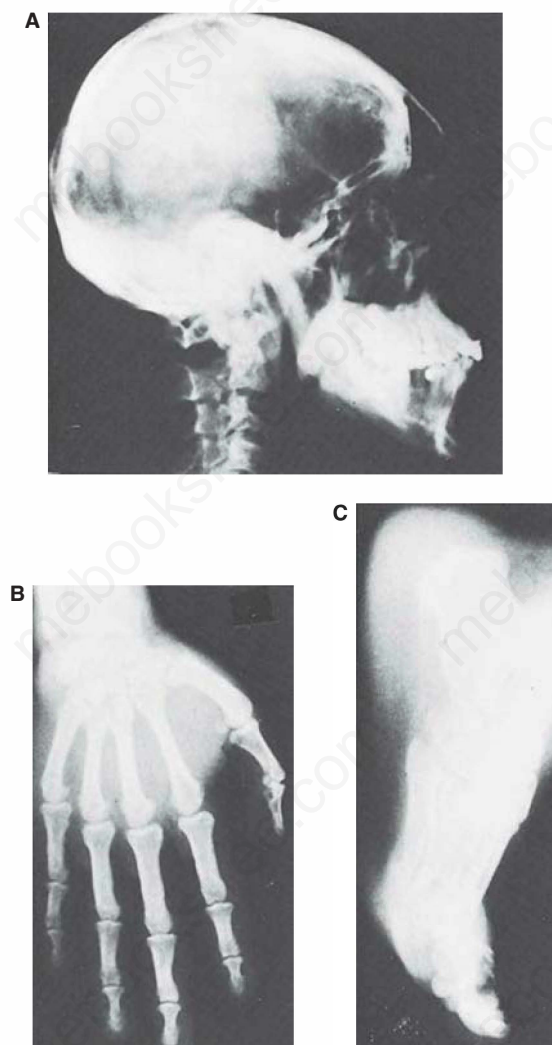
FIGURE 4–20 Radiologic signs in acromegaly. **A**: Skull with enlarged sella turcica and frontal sinuses, thickening of the calvarium, and enlargement of the mandible. **B**: Hand with enlarged sesamoid bone and increased soft tissue shadows. **C**: Thickened heel pad. (Reproduced with permission from Levin SR. Manifestations and treatment of acromegaly. *Calif Med*. 1972 Mar;116(3):57-64.)

1 ng/mL (47 pmol/L), and this lack of response establishes the diagnosis. Supersensitive GH assays have been developed and are becoming commercially available. With these assays, normal individuals may suppress GH levels to less than 0.1 ng/mL. Thus, the criteria expressed earlier may need to be adjusted in the near future (to <0.04 ng/mL). Since many assays lack sufficient accuracy; however, a cutoff of less than 0.1 ng/mL is still recommended.

B. IGF-1 measurement   Measurement of IGF-1 (see Chapter 6) is a useful means of establishing the diagnosis of GH hypersecretion. IGF-1 results must be interpreted according to age- and sex-adjusted normative data. IGF-1 levels directly reflect GH activity. IGF-1 has a long half-life, so that IGF-1 levels fluctuate much less than GH levels. IGF-1 levels are

elevated in virtually all patients with acromegaly (normal ranges vary widely in different laboratories, and some commercial assays are not reliable).

C. Tumor localization Radiographic localization of the pituitary adenoma causing acromegaly is usually straightforward (see Neuroradiologic Evaluation, earlier). In virtually all patients, tumor location and size can be shown by MRI; 90% have tumors greater than 1 cm in diameter that are readily visualized. In the rare patient with normal neuroradiologic studies, an extrapituitary ectopic source of GH or GHRH should be considered. If scans suggest diffuse pituitary enlargement or hyperplasia, ectopic GHRH should also be suspected.

Differential Diagnosis

A. Other causes of GH or IGF-1 hypersecretion The presence of clinical features of GH excess, elevated GH and IGF-1 secretion, and abnormal GH dynamics, together with the demonstration of a pituitary tumor by neuroradiologic studies, are diagnostic of acromegaly. However, other conditions associated with GH hypersecretion must be considered in the differential diagnosis. These include anxiety, exercise, acute illness, chronic renal failure, cirrhosis, starvation, protein-calorie malnutrition, anorexia nervosa, and type 1 (insulin-dependent) diabetes mellitus. Estrogen therapy may increase GH responsiveness to various stimuli. These conditions may be associated with abnormal GH suppressibility by glucose and by abnormal GH responsiveness to TRH; however, patients with these conditions do not have clinical manifestations of GH excess and are thus readily differentiated from patients with acromegaly. In addition, the conditions listed earlier do not lead to elevation of IGF-1 concentrations. Use of testosterone or the contraceptive depot medroxyprogesterone acetate has been associated with modest elevations in IGF-1.

B. Ectopic GH or GHRH secretion These rare patients with acromegaly due to ectopic secretion of GH or GHRH have typical clinical manifestations of acromegaly. This may occur in lung carcinoma, carcinoid tumors, and pancreatic islet cell tumors. These syndromes should be suspected in patients with a known extrapituitary tumor who have GH excess or in those with clinical and biochemical features of acromegaly who have radiologic procedures that show normal pituitary glands or that suggest diffuse pituitary enlargement or hyperplasia.

Treatment

All patients with acromegaly should undergo therapy to halt progression of the disorder and to prevent late complications and excess mortality. The objectives of therapy are removal or destruction of the pituitary tumor, reversal of GH hypersecretion, and maintenance of normal anterior and posterior pituitary function. These objectives are currently attainable in most patients, especially those with smaller tumors and only moderate GH hypersecretion. In patients with large tumors who have marked GH hypersecretion, several therapies are usually required to achieve normal GH secretion.

The criteria for an adequate response to therapy continue to evolve. Some reports have described increased late mortality in patients with GH levels by radioimmunoassay greater than 2.5 ng/mL (116 pmol/L) after therapy. Therefore, current guidelines for remission are a fasting GH of 1 ng/mL (47 pmol/L) or less and a glucose-suppressed GH of 1 ng/mL (47 pmol/L) or less accompanied by a normal level of IGF-1.

The initial therapy of choice is transsphenoidal microsurgery because of its high success rate, rapid reduction of GH levels, the low incidence of postoperative hypopituitarism, and the low surgical morbidity rate. Patients with persisting GH hypersecretion after surgery should be treated medically with somatostatin analogs, dopamine agonists, or a GH receptor antagonist. Radiation therapy should be reserved for those patients with inadequate responses to surgery and medical therapy.

A. Surgical treatment Transsphenoidal selective adenoma removal is the procedure of choice; craniotomy is necessary in the rare patient in whom major suprasellar extension precludes the transsphenoidal approach. Successful reduction of GH levels is achieved in approximately 60% to 80% of patients. In those with small or moderate-sized tumors (<2 cm), success is achieved in more than 80%, whereas in those with larger tumors and basal GH levels greater than 50 ng/mL (2325 pmol/L)—and particularly in those with major extrasellar extension of the adenoma—successful responses occur in only 30% to 60%. Recurrence rates in those with a successful initial response are low (about 5% of patients at our institution). Surgical complications (discussed earlier) occur in less than 2%.

B. Medical treatment Octreotide acetate, a somatostatin analog, was the first effective medical therapy for patients with acromegaly. However, the drug required high doses (100-500 µg) given subcutaneously three times daily. Its use in acromegaly has been superseded by sustained-release somatostatin analogs with activities lasting up to 1 month. Preparations include octreotide LAR, lanreotide acetate, and pasireotide diaspartate given by injection every 4 weeks. Octreotide LAR normalizes GH and IGF-1 levels in 75% of patients when used at doses of 20 to 40 mg/mo; however, tumor reduction occurs in a much smaller percentage. Pasireotide binds with high affinity to 4 of the 5 somatostatin receptor subtypes and may be more effective than octreotide at normalizing GH and IGF-1 levels. The long-acting agents have become the therapy of choice for patients with residual GH hypersecretion following surgery. Side effects of this class of agents consist mainly of gastrointestinal symptoms and the development of gallstones. Impaired glucose tolerance has been reported, especially with pasireotide.

The dopamine agonist cabergoline normalizes IGF-1 levels in about 30% of acromegalic patients when used at doses of 1 to 2 mg/wk. However, it has not been commonly used as sole therapy. When cabergoline is added to somatostatin analog therapy, the number of patients with normalization of GH and IGF-1 levels is increased.

Pegvisomant is a GH receptor antagonist that has been shown to be an effective therapy for the treatment of acromegaly. In doses of 10 to 20 mg/d given by subcutaneous injection, it reduces IGF-1 levels to normal in more than 90% of patients. Although there are concerns regarding the continued excess GH secretion and possible tumor progression, serious problems have not arisen to date. Although pegvisomant has very good efficacy, its use may be limited by its high cost and the need for daily injections. Accordingly, it is currently used mainly in those patients who have failed therapy with surgery and somatostatin analogs.

C. Radiotherapy   Conventional supervoltage irradiation in doses of 4500 to 5000 cGy, although ultimately successful in 60% to 80% of patients, should not be used, because GH levels may not return to normal until 10 to 15 years after therapy. In one series, GH levels were under 10 ng/mL (460 pmol/L) in only 38% of patients at 2 years posttreatment; however, at 5 and 10 years, 73% and 81% had achieved such levels. The incidence of hypopituitarism is appreciable, and in this series hypothyroidism occurred in 19%, hypoadrenalism in 38%, and hypogonadism in approximately 50% to 60% of patients as a consequence of radiotherapy. Gamma knife radiosurgery has also been used for tumors confined to the sella. Current series, although limited, suggest remission rates of about 50% to 70% at 2 years following therapy.

Response to Treatment

In patients with successful reduction in GH hypersecretion, there is cessation of bone overgrowth. In addition, these patients experience considerable clinical improvement, including reduction in soft tissue bulk of the extremities, decreased facial puffiness, increased energy, and cessation of hyperhidrosis, heat intolerance, and oily skin. Bony changes typically do not regress. Headache, carpal tunnel syndrome, arthralgias, sleep apnea, and photophobia are also reversible with successful therapy. Glucose intolerance and hyperinsulinemia as well as hypercalciuria are also reversed in most cases. Excess mortality associated with acromegaly can be reversed if GH levels are normalized.

Posttreatment Follow-Up

Posttreatment assessment includes evaluation of GH secretion, anterior pituitary function, and tumor size. Patients undergoing surgery should postoperative assessment of pituitary function. Measurement of IGF-1 and GH should be performed at 12 weeks or later. If GH is >1 ng/mL, measure nadir GH level after a glucose load. Those with persistent GH hypersecretion (>1 ng/mL [47 pmol/L]) should receive further therapy with somatostatin analogs. Patients with postoperative GH levels under 1 ng/mL (47 pmol/L) should have follow-up GH and IGF-1 determinations at 6-month intervals for 2 years and yearly thereafter to rule out recurrences. Recurrent elevations in IGF-1 should prompt a repeat MRI of the sella. Late hypopituitarism after surgery alone does not occur. GH replacement therapy in patients with history of acromegaly and hypopituitarism is controversial and further studies are needed.

3. ACTH-SECRETING PITUITARY ADENOMAS: CUSHING DISEASE

In 1932, Harvey Cushing documented the presence of small basophilic pituitary adenomas in six of eight patients with clinical features of adrenocortical hyperfunction. Years later, ACTH hypersecretion was identified from such tumors and found to be the cause of bilateral adrenal hyperplasia. Pituitary ACTH hypersecretion (Cushing disease) is now recognized as the most common cause of endogenous hypercortisolism (Cushing syndrome) and must be distinguished from the other forms of adrenocorticosteroid excess—ectopic ACTH syndrome and adrenal tumors (see Chapter 9).

Pathology

ACTH-secreting pituitary tumors exist in virtually all patients with Cushing disease. These tumors are usually benign microadenomas under 10 mm in diameter; 50% are 5 mm or less in diameter, and microadenomas as small as 1 mm have been described. These tumors in Cushing disease are either basophilic or chromophobe adenomas and may be found anywhere within the anterior pituitary. Rarely, ACTH-secreting tumors are large, with invasive tendencies, and malignant tumors have rarely been reported.

Histologically, the tumors are composed of compact sheets of uniform, well-granulated cells (granule size, 200-700 nm by electron microscopy) with a sinusoidal arrangement and a high content of ACTH and its related peptides (β-LPH, β-endorphin). A zone of perinuclear hyalinization (Crooke changes) is frequently observed as a result of exposure of the corticotroph cells to prolonged hypercortisolism (see Figure 4–7). A specific ultrastructural finding in these adenomas is the deposition of bundles of perinuclear microfilaments that encircle the nucleus; these are the ultrastructural equivalent of Crooke hyaline changes seen on light microscopy. In contrast to the adenoma's cells, ACTH content in the portion of the anterior pituitary not involved with the tumor is decreased.

Diffuse hyperplasia of anterior pituitary corticotrophs or adenomatous hyperplasia, presumed to result from hypersecretion of CRH, occurs rarely.

The adrenal glands in Cushing disease are enlarged, weighing 12 to 24 g (normal, 8-10 g). Microscopic examination shows a thickened cortex due to hyperplasia of both the zona reticularis and zona fasciculata; the zona glomerulosa is normal. In some cases, ACTH-secreting pituitary adenomas cause bilateral nodular hyperplasia; the adrenals show diffuse bilateral cortical hyperplasia and the presence of one or more nodules that vary from microscopic to several centimeters in diameter, with multiple small nodules being the most common.

Pathogenesis

The weight of current evidence is that Cushing disease is a primary pituitary disorder and that hypothalamic abnormalities are secondary to hypercortisolism. The endocrine abnormalities in Cushing disease are as follows: (1) hypersecretion of ACTH, with bilateral adrenocortical hyperplasia and hypercortisolism; (2) absent circadian periodicity of ACTH and cortisol secretion;

(3) absent responsiveness of ACTH and cortisol to stress (hypoglycemia or surgery); (4) abnormal negative feedback of ACTH secretion by glucocorticoids; and (5) subnormal responsiveness of GH, TSH, and gonadotropins to stimulation.

Evidence that Cushing disease is a primary pituitary disorder is based on the high frequency of pituitary adenomas, the response to their removal, and the interpretation of hypothalamic abnormalities as being secondary to hypercortisolism. In addition, molecular studies have found that nearly all corticotroph adenomas are monoclonal. The ubiquitin-specific protease 8 gene (*USP8*) has been found to be mutated in some corticotroph adenomas. The product of this gene has deubiquitinase activity and appears to be involved in the regulation of epidermal growth factor receptor (EGFR) recycling. These findings suggest that ACTH hypersecretion arises from a spontaneously developing pituitary adenoma and that the resulting hypercortisolism suppresses the normal hypothalamic-pituitary axis and CRH release, thereby abolishing the hypothalamic regulation of circadian variability and stress responsiveness.

Analysis of the response to therapy by pituitary microsurgery sheds some light on the pathogenesis of Cushing disease. Selective removal of pituitary microadenomas by transsphenoidal microsurgery corrects ACTH hypersecretion and hypercortisolism in most patients. After selective removal of the pituitary adenoma, the following return to normal: the circadian rhythmicity of ACTH and cortisol, the responsiveness of the hypothalamic-pituitary axis to hypoglycemic stress, and the dexamethasone suppressibility of cortisol secretion.

Clinical Features

Cushing disease presents with the signs and symptoms of hypercortisolism and adrenal androgen excess (see Chapter 9). The onset of these features is usually insidious, developing over months or years. Obesity (with predominantly central fat distribution), hypertension, glucose intolerance, and gonadal dysfunction (amenorrhea or impotence) are common features. Other common manifestations include moon (rounded) facies, plethora, osteopenia, proximal muscle weakness, easy bruisability, psychologic disturbances, violaceous striae, hirsutism, acne, poor wound healing, and superficial fungal infections. Unlike patients with the classic form of ectopic ACTH syndrome, patients with Cushing disease rarely have hypokalemia, weight loss, anemia, or hyperpigmentation. Virilization, observed occasionally in patients with adrenal carcinoma, is unusual in Cushing disease. Clinical symptoms related to the ACTH-secreting primary tumor itself, such as headache or visual impairment, are rare because of the small size of these adenomas.

The usual age range is 20 to 40 years, but Cushing disease has been reported in infants and in patients over 70 years. There is a female:male ratio of approximately 8:1. In contrast, the ectopic ACTH syndrome occurs more commonly in men (male:female ratio of 3:1).

Diagnosis

The initial step in the diagnosis of an ACTH-secreting pituitary adenoma is the documentation of endogenous hypercortisolism, which is confirmed by increased urinary free cortisol secretion, abnormal cortisol suppressibility to low-dose dexamethasone, and/or abnormal late night salivary cortisol measurement. The differentiation of an ACTH-secreting pituitary tumor from other causes of hypercortisolism must be based on biochemical studies, including the measurement of basal plasma ACTH levels and central venous sampling, to detect a central to peripheral gradient of ACTH levels (see Chapter 9). Imaging of the sella with MRI may help to identify a pituitary adenoma in only 50% to 70% of patients with ACTH-dependent hypercortisolism. The diagnosis and differential diagnosis of Cushing syndrome are presented in Chapter 9.

Treatment

Transsphenoidal microsurgery is the procedure of choice in Cushing disease. A variety of other therapies—operative, radiologic, pharmacologic—are discussed later.

A. Surgical treatment Selective transsphenoidal resection of ACTH-secreting pituitary adenomas is the initial treatment of choice. At operation, meticulous exploration of the intrasellar contents by an experienced neurosurgeon is required. The tumor, which is usually found within the anterior lobe tissue, is selectively removed, and normal gland is left intact.

In about 85% of patients with microadenomas, selective microsurgery is successful in correcting hypercortisolism. Surgical damage to anterior pituitary function is rare, but most patients develop transient secondary adrenocortical insufficiency requiring postoperative glucocorticoid support until the hypothalamic-pituitary-adrenal axis recovers, usually in 2 to 18 months. The role of total hypophysectomy is currently unclear; however, hemihypophysectomy based on central venous sampling lateralization is successful in only about 50% to 70% of patients.

By contrast, transsphenoidal surgery is successful in only 25% of the 10% to 15% of patients with Cushing disease with pituitary macroadenomas or in those with extrasellar extension of tumor.

Transient diabetes insipidus occurs in about 10% of patients, but other surgical complications (eg, hemorrhage, cerebrospinal fluid rhinorrhea, infection, visual impairment, permanent diabetes insipidus) are rare. Hypopituitarism occurs only in patients who undergo total hypophysectomy.

Before the introduction of pituitary microsurgery, bilateral total adrenalectomy was the preferred treatment of Cushing disease and may still be employed in patients in whom other therapies are unsuccessful. Total adrenalectomy, which can now be performed laparoscopically, corrects hypercortisolism but produces permanent hypoadrenalism, requiring lifelong glucocorticoid and mineralocorticoid therapy. The ACTH-secreting pituitary adenoma persists and may progress, causing hyperpigmentation and invasive complications (Nelson syndrome; see later). Persistent hypercortisolism may occasionally follow total adrenalectomy as ACTH hypersecretion stimulates adrenal remnants or congenital rests.

B. Radiotherapy Conventional radiotherapy of the pituitary is of benefit in patients who have persistent or recurrent

disease following pituitary microsurgery. In these patients, reported remission rates are 55% to 70% at 1 to 3 years after radiotherapy.

Gamma knife radiosurgery achieves remission rates of 65% to 75%. However, as noted earlier, both forms of radiotherapy cause late loss of pituitary function in more than 50% of patients and visual deficits can occur with damage to the optic chiasm or cranial nerves.

C. Medical treatment Drugs that inhibit adrenal cortisol secretion are useful in Cushing disease, often as adjunctive therapy (see Chapter 9).

Ketoconazole, an imidazole derivative, has been found to inhibit adrenal steroid biosynthesis. It inhibits the cytochrome P450 enzymes P450scc and P450c11. In daily doses of 600 to 1200 mg, ketoconazole has been effective in the management of Cushing syndrome. Hepatotoxicity is common, however, but may be transient. **Metyrapone**, which inhibits P450c11, has also been utilized to reduce cortisol hypersecretion. The use of these drugs is accompanied by increased ACTH levels that may overcome the enzyme inhibition. In addition, they cause gastrointestinal side effects that may limit their effectiveness. More effective control of hypercortisolism with fewer side effects is obtained by combined use of these agents. Adequate data are not available on the long-term use of these drugs as the sole treatment of Cushing disease. Thus, ketoconazole and metyrapone are generally used while awaiting a response to therapy or in the preparation of patients for surgery.

The adrenolytic drug **mitotane** results in adrenal atrophy predominantly of the zonae fasciculata and reticularis. Remission of hypercortisolism is achieved in approximately 80% of patients with Cushing disease, but most relapse after therapy is discontinued. Mitotane therapy is limited by the delayed response, which may take weeks or months, and by the frequent side effects, including severe nausea, vomiting, diarrhea, somnolence, and skin rash.

The anesthesia induction agent **etomidate** inhibits P450c11. It must be given intravenously and is generally reserved for life-threatening cases of hypercortisolism resistant to conventional therapy or when oral medications are contraindicated.

Pituitary-directed therapies include **cabergoline** and **pasireotide diaspartate**. The type 2 dopamine (D2) receptor is present on the majority of corticotroph adenomas. Doses of cabergoline of 1 to 7 mg per week has been shown to be effective at normalizing urinary free cortisol levels in 10% to 30% of patients, can be given orally, and is generally well tolerated. In clinical practice, its role is in a minority of patients with mild hypercortisolism. Pasireotide is a somatostatin (sst) analog that binds with high affinity to sst_1, sst_2, sst_3, and especially sst_5 subtypes of the sst receptor which are present on corticotroph tumors. It was approved for clinical use at doses of 0.6 to 0.9 mg subcutaneously twice daily. In one study, urinary free cortisol normalized in 20% of patients and remained controlled at 12 months. In patients who do not respond after 3 months of therapy despite titration to 0.9 mg twice daily, it should be discontinued or combined with another therapy. Like other sst analogs, adverse effects include gallstones and gastrointestinal

symptoms. Hyperglycemia appears to be more common, however, occurring in a majority of patients. Over half required glucose-lowering therapy in one study. **Mifepristone** is a glucocorticoid-receptor antagonist approved for treatment of hyperglycemia in patients with Cushing syndrome in the United States. It is best known for its anti-progestational properties (RU-486) and has been used for the treatment of endometriosis and uterine myomas and to induce abortions. It has been found to act as a glucocorticoid receptor antagonist at higher doses. It is typically started at 300 mg daily and titrated up to 1200 mg daily. The dose adjustment is based on metabolic parameters (eg, glucose) given that ACTH and cortisol levels increase as a result of glucocorticoid receptor blockade. Side effects include adrenal insufficiency, fatigue, nausea, headache, hypokalemia, edema, and endometrial thickening in women. Given that cortisol levels are not low, a high degree of clinical suspicion is necessary to identify adrenal insufficiency in patients using mifepristone. A dexamethasone dose of 2 to 4 mg daily for every 300 mg of mifepristone is recommended in patients with suspected adrenal insufficiency associated with glucocorticoid receptor antagonism.

4. NELSON SYNDROME

The clinical appearance of an ACTH-secreting pituitary adenoma following bilateral adrenalectomy as initial therapy for Cushing disease was first described by Nelson and coworkers in 1958. However, with the evolution of pituitary microsurgery as the initial therapy for Cushing disease, the Nelson syndrome is now an unusual occurrence.

Pathogenesis

It now seems likely that Nelson syndrome represents the clinical progression of a preexisting adenoma after the restraint of hypercortisolism on ACTH secretion and tumor growth is removed. Thus, following adrenalectomy, the suppressive effect of cortisol is no longer present, ACTH secretion increases, and the pituitary adenoma may progress.

Incidence

Prior to the development of transsphenoidal surgery, when bilateral adrenalectomy was the initial therapy for Cushing disease, the incidence of Nelson syndrome ranged from 10% to 78% depending on what criteria were used for diagnosis (see Chapter 9). In one series, approximately 30% of patients adrenalectomized for Cushing disease developed classic Nelson syndrome with progressive hyperpigmentation and an obvious ACTH-secreting tumor; another 50% developed evidence of a microadenoma without marked progression; and about 20% never developed a progressive tumor. The reasons for these differences in clinical behavior are uncertain. Prophylactic pituitary radiotherapy is controversial. At present, when adrenalectomy is utilized only in those patients who fail pituitary microsurgery, the incidence of Nelson syndrome is less than 10%. Nevertheless, continued examination, including plasma ACTH levels and MRI, is required following bilateral adrenalectomy in patients with Cushing disease.

Clinical Features

The pituitary tumors in patients with classic Nelson syndrome can be among the most aggressive and rapidly growing of all pituitary tumors. These patients present with hyperpigmentation and with manifestations of an expanding intrasellar mass lesion. Visual field defects, headache, cavernous sinus invasion with extraocular muscle palsies, and even malignant changes with local or distant metastases may occur. Pituitary apoplexy may also complicate the course of these tumors.

Diagnosis

Plasma ACTH levels are markedly elevated, usually over 1000 pg/mL (222 pmol/L) and often as high as 10,000 pg/mL (2220 pmol/L). MRI defines the extent of the tumor.

Treatment

Pituitary surgery by the transsphenoidal approach is the initial mode of treatment. Complete resection is usually not possible, because of the large size of these tumors. Conventional radiotherapy or gamma knife radiosurgery is employed postoperatively in patients with residual tumor or extrasellar extension.

5. THYROTROPIN-SECRETING PITUITARY ADENOMAS

TSH-secreting pituitary adenomas are rare tumors manifested as hyperthyroidism with goiter in the presence of elevated TSH. Patients with TSH-secreting tumors are often resistant to routine ablative thyroid therapy, requiring large, often multiple doses of ^{131}I and several operations for control of thyrotoxicosis. Histologically, the tumors are chromophobe adenomas. They are often very large and cause visual impairment, which alerts the physician to a pituitary abnormality. Patients with these tumors do not have extrathyroidal systemic manifestations of Graves disease such as ophthalmopathy or dermopathy.

The diagnosis is based on findings of hyperthyroidism (elevated T_4 or T_3) with elevated serum TSH and alpha subunit, and neuroradiologic studies consistent with pituitary tumor. The differential diagnosis includes those patients with primary hypothyroidism (thyroid failure) who develop major hyperplasia of pituitary thyrotrophs and lactotrophs with sellar enlargement and occasional suprasellar extension. These patients can be identified by symptoms of hypothyroidism and low thyroid hormone levels. Thyroid hormone resistance may be more difficult to exclude as T_4 and T_3 levels can also be elevated. These patients, however, tend to be clinically euthyroid or hypothyroid, have a family history of thyroid hormone resistance, and generally do not present with sellar masses. Alpha subunit levels in these patients are usually not elevated.

Treatment should be directed initially at the adenoma via the transsphenoidal microsurgical approach. However, additional therapy is usually required because of the large size of these adenomas.

Somatostatin analogs normalize TSH and T_4 levels in more than 70% of these patients when given in doses similar to those used for the treatment of acromegaly (see discussion earlier). Shrinkage of the tumor has been observed in about 40% of patients.

If tumor growth and TSH hypersecretion cannot be controlled by surgery and somatostatin analogs, the next step is pituitary irradiation. In addition, such patients may also require ablative therapy of the thyroid with either ^{131}I or surgery to control their thyrotoxicosis.

6. GONADOTROPIN-SECRETING PITUITARY ADENOMAS

Although many pituitary adenomas synthesize gonadotropins (especially FSH) and their subunits, only a minority of these patients has elevated serum levels of FSH or LH. The majority of these tumors produce FSH and the alpha subunit, but tumors secreting both FSH and LH and a tumor secreting only LH have been described.

Gonadotropin-secreting pituitary adenomas are usually large chromophobe adenomas presenting with visual impairment. Most patients have hypogonadism and many have panhypopituitarism. Hormonal evaluation reveals elevated FSH in some patients accompanied by normal LH values. Basal levels of the alpha subunit may also be elevated. The presence of elevation of both FSH and LH should suggest primary hypogonadism. TRH stimulation leads to an increase in FSH secretion in 33% and an increase in LH-β in 66% of patients.

Therapy for gonadotropin-secreting adenomas has been directed at surgical removal. Because of their large size, adequate control of the tumor is difficult to achieve, and radiotherapy is usually required.

7. ALPHA SUBUNIT-SECRETING PITUITARY ADENOMAS

Excessive quantities of the alpha subunit of the glycoprotein pituitary hormones have been observed in association with the hypersecretion of many anterior pituitary hormones (TSH, GH, PRL, LH, FSH). However, pure alpha subunit hypersecretion has been identified in several patients with large invasive chromophobe adenomas and partial panhypopituitarism. Thus, the determination of the alpha subunit may be a useful marker in patients with presumed "nonfunctioning" pituitary adenomas.

8. NONFUNCTIONAL PITUITARY ADENOMAS

"Nonfunctional" chromophobe adenomas once represented approximately 80% of all primary pituitary tumors; however, with clinical application of sensitive immunoassays of anterior pituitary hormones, these tumors currently account for a minority of all pituitary adenomas and many exhibit histochemical staining for hormones. Thus, the great majority of these chromophobe adenomas have now been documented to be PRL-secreting; a smaller number secrete TSH or the gonadotropins.

Nonfunctional tumors are usually large when the diagnosis is established; headache and visual field defects are the usual presenting symptoms. However, endocrine manifestations are usually present for months to years before the diagnosis is made, with gonadotropin deficiency being the most common initial symptom. Hypothyroidism and hypoadrenalism are also common, but the symptoms are subtle and may be missed.

Evaluation should include MRI and visual field testing; endocrine studies should include assessment of pituitary hormones and end-organ function to determine whether the adenoma is hypersecreting or whether hormonal replacement is needed.

Because these tumors are generally large, both surgery and radiation therapy are usually required to prevent tumor progression or recurrence. In the absence of an endocrine index of tumor hypersecretion such as PRL excess, serial scans at yearly intervals are required to assess the response to therapy and to detect possible recurrence.

9. PITUITARY CARCINOMA

Pituitary carcinomas are extremely rare with fewer than 200 cases reported and defined by the distant metastases of a pituitary tumor. Most present as hormone-producing invasive macroadenomas with symptoms of mass effect. ACTH- and PRL-secreting tumors are most common. Metastases may present many years after diagnosis of the primary pituitary tumor. Survival rates are low, and treatment may include additional surgery, radiotherapy, or chemotherapy (eg, temozolomide).

REFERENCES

Neuroendocrinology

Anderson JR, Antoun N, Burnet N, et al. Neurology of the pituitary gland. *J Neurol Neurosurg Psychiat*. 1999;66:703. [PMID: 10329742]

Ben-Jonathan N, Hnasko R. Dopamine as a prolactin (PRL) inhibitor. *Endocr Rev*. 2001;22:724. [PMID: 11739329]

Burrows HL, Douglas KR, Seasholtz AF, Camper SA. Genealogy of the anterior pituitary gland: tracing a family tree. *Trends Endocrinol Metab*. 1999;10:343. [PMID: 10511693]

Coppari R, Ramadori G, Elmquist JK. The role of transcriptional regulators in central control of appetite and body weight. *Nat Clin Pract Endocrinol Metab*. 2009;5:160. [PMID: 19229236]

Davies T, Marians R, Latif R. The TSH receptor reveals itself. *J Clin Invest*. 2002;110:209. [PMID: 12122107]

Elmquist JK. Hypothalamic pathways underlying the endocrine, autonomic, and behavioral effects of leptin. *Int J Obes Relat Metab Disord*. 2001;25:S78. [PMID: 11840221]

Fauquier T, Lacampagne A, Travo P, Bauer K, Mollard P. Hidden face of the anterior pituitary. *Trends Endocrinol Metab*. 2002;13:304. [PMID: 12163233]

Fourman LT, Fazeli PK. Neuroendocrine causes of amenorrhea—an update. *J Clin Endocrinol Metab*. 2015;100:812. [PMID: 25581597]

Gudin JA, Laitman A, Nalamachu S. Opioid related endocrinopathy. *Pain Med*. 2015;16(Suppl 1):S9. [PMID: 26461076]

Keenan DM, Veldhuis JD. Pulsatility of hypothalamo-pituitary hormones: a challenge in quantification. *Physiology* (Bethesda). 2016;31:34. [PMID: 26674550]

Le Roith D, Bondy C, Yakar S, Liu JL, Butler A. The somatomedin hypothesis. *Endocr Rev*. 2001;22:53. [PMID: 11159816]

Moller M, Baeres FM. The anatomy and innervation of the mammalian pineal gland. *Cell Tissue Res*. 2002;309:139. [PMID: 12111544]

Muccioli G, Tschop M, Papotti M, Deghenghi R, Heiman M, Ghigo E. Neuroendocrine and peripheral activities of ghrelin: implications in metabolism and obesity. *Eur J Pharmacol*. 2002;440:235. [PMID: 12007539]

Richard D, Lin Q, Timofeeva E. The corticotropin-releasing factor family of peptides and CRF receptors: their roles in the regulation of energy balance. *Eur J Pharmacol*. 2002;12:189. [PMID: 12007535]

Tasker JG. Rapid glucocorticoid actions in the hypothalamus as a mechanism of homeostatic integration. *Obesity*. 2006;14:259S. [PMID: 17021378]

Zigman JM, Elmquist JK. Minireview: from anorexia to obesity—the yin and yang of body weight control. *Endocrinology*. 2003;144:3749. [PMID: 12933644]

Pituitary Function Testing and Neuroradiology

Biller BM, Samuels MH, Zagar A, et al. Sensitivity and specificity of six tests for the diagnosis of adult GH deficiency. *J Clin Endocrinol Metab*. 2002;87:2067. [PMID: 11994342]

Gasco V, Corneli G, Rovere S, et al. Diagnosis of adult GH deficiency. *Pituitary*. 2008;11:121. [PMID: 18404387]

Kazlauskaite R, Evans AT, Villabona CV, et al. Corticotropin tests for hypothalamic-pituitary-adrenal insufficiency: a metaanalysis. *J Clin Endocrinol Metab*. 2008;93:4245. [PMID: 18697868]

Patronas N, Bulakbasi N, Stratakis CA, et al. Spoiled gradient recalled acquisition in the steady state technique is superior to conventional postcontrast spin echo technique for magnetic resonance imaging detection of adrenocorticotropin-secreting pituitary tumors. *J Clin Endocrinol Metab*. 2003;88:1565. [PMID: 12679440]

Raff H. Utility of salivary cortisol measurements in Cushing's syndrome and adrenal insufficiency. *J Clin Endocrinol Metab*. 2009;94:3647. [PMID: 19602555]

Smith TP, Kavanagh L, Healy ML, McKenna TJ. Technology insight: measuring prolactin in clinical samples. *Nat Clin Pract Endocrinol Metab*. 2007;3:279. [PMID: 17315036]

Pituitary Adenomas: General

Asa SL, Ezzat S. The pathogenesis of pituitary tumors. *Annu Rev Pathol*. 2009; 4:97. [PMID: 19400692]

Brada M, Jankowska P. Radiotherapy for pituitary adenomas. *Endocrinol Metab Clin N Am*. 2008;37:263. [PMID: 18226740]

Dekkers OM, Pereira AM, Romijn JA. Treatment and follow-up of clinically nonfunctioning pituitary macroadenomas. *J Clin Endocrinol Metab*. 2008;93:3717. [PMID: 18682516]

Elston MS, McDonald KL, Clifton-Bligh RJ, Robinson BG. Familial pituitary tumor syndromes. *Nat Rev Endocrinol*. 2009;5:453. [PMID: 19564887]

Fernandez A, Karavitaki N, Wass JA. Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK). *Clin Endocrinol (Oxf)*. 2010;72:377. [PMID: 19650784]

Jagannathan J, Kanter AS, Sheehan JP, Jane JA Jr, Laws ER Jr. Benign brain tumors: sellar/parasellar tumors. *Neurol Clin*. 2007;25:1231. [PMID: 17964033]

Komninos J, Vlassopoulou V, Protopapa D, et al. Tumors metastatic to the pituitary gland: case report and literature review. *J Clin Endocrinol Metab*. 2004;89:574. [PMID: 14764764]

Krikorian A, Aron D. Evaluation and management of pituitary incidentalomas—revisiting an acquaintance. *Nat Clin Pract Endocrinol Metab*. 2006;2:138. [PMID: 16932273]

Lopes MB, Scheithauer BW, Schiff D. Pituitary carcinoma: diagnosis and treatment. *Endocrine*. 2005;28:115. [PMID: 16311418]

ACTH: Cushing Disease

Assié G, Bahurel H, Bertherat J, et al. The Nelson's syndrome… revisited. *Pituitary*. 2004;7:209. [PMID: 16132203]

Creemers SG, Hofland LJ, Lamberts SW, Feelders RA. Cushing's syndrome: an update on current pharmacotherapy and future directions. *Expert Opin Pharmacother*. 2015;16:1829. [PMID: 26133755]

Findling JW, Raff H. Cushing's syndrome: important issues in diagnosis and management. *J Clin Endocrinol Metab*. 2006;91:3746. [PMID: 16868050]

Jehle S, Walsh JE, Freda PU, Post KD. Selective use of bilateral inferior petrosal sinus sampling in patients with adrenocorticotropin-dependent Cushing's syndrome prior to transsphenoidal surgery. *J Clin Endocrinol Metab*. 2008;93:4624. [PMID: 18796519]

Lacroix A, Feelders RA, Stratakis CA, Nieman LK. Cushing's syndrome. *Lancet*. 2015;386:913. [PMID: 26004339]

Nieman LK, Biller BM, Findling JW, et al. The diagnosis of Cushing's syndrome: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2008;93:1526. [PMID:18334580]

Sbiera S, Deutschbein T, Weigand I, Reincke M, Fassnacht M, Allolio B. The new molecular landscape of Cushing's disease. *Trends Endocrinol Metab*. 2015;26:573. [PMID: 26412158]

Vance ML. Cushing's disease: radiation therapy. *Pituitary*. 2009;12:11. [PMID: 18437577]

Growth Hormone: Acromegaly

Freda PU, Post KD, Powell JS, Wardlaw SL. Evaluations of disease status with sensitive measures of growth hormone secretion in 60 postoperative patients with acromegaly. *J Clin Endocrinol Metab*. 1998;83:3808. [PMID: 9814451]

Melmed S. Acromegaly pathogenesis and treatment. *J Clin Invest*. 2009;119:3189. [PMID: 19884662]

PRL: Prolactinoma

Gibney J, Smith TP, McKenna TJ. Clinical relevance of macroprolactin. *Clin Endocrinol (Oxf)*. 2005;62:633. [PMID: 15943822]

Molitch ME. Medication-induced hyperprolactinemia. *Mayo Clin Proc*. 2005;80:1050. [PMID: 16092584]

Schade R, Andersohn F, Suissa S, Haverkamp W, Garbe E. Dopamine agonists and the risk of cardiac-valve regurgitation. *N Engl J Med*. 2007;356:29. [PMID: 17202453]

Schlechte JA. Long-term management of prolactinomas. *J Clin Endocrinol Metab*. 2007;92:2861. [PMID: 17682084]

Gonadotropins (LH/FSH): Gonadotropin-Secreting Pituitary Tumors

Ntali G, Capatina C, Grossman A, Karavitaki N. Clinical review: functioning gonadotroph adenomas. *J Clin Endocrinol Metab*. 2014;99:4423. [PMID: 25166722]

Shomali ME, Katznelson L. Medical therapy for gonadotroph and thyrotroph tumors. *Endocrinol Metab Clin North Am*. 1999;28:223. [PMID: 10207693]

TSH-Secreting Pituitary Tumors

Brucker-Davis F, Oldfield EH, Skarulis MC, Doppman JL, Weintraub BD. Thyrotropin-secreting pituitary tumors: diagnostic criteria, thyroid hormone sensitivity, and treatment outcome in 25 patients followed at the National Institutes of Health. *J Clin Endocrinol Metab*. 1999;84:476. [PMID: 10022404]

Laws ER, Vance ML, Jane JA, Jr. TSH adenomas. *Pituitary*. 2006:9:313. [PMID: 17080266]

Hypopituitarism and Other Hypothalamic-Pituitary Disorders

Bihan H, Christozova V, Dumas JL, et al. Sarcoidosis: clinical, hormonal, and magnetic resonance imaging (MRI) manifestations of hypothalamic-pituitary disease in 9 patients and review of the literature. *Medicine*. 2007;86:259. [PMID: 17873755]

Carpinteri R, Patelli I, Casanueva FF, Giustina A. Pituitary tumours: inflammatory and granulomatous expansive lesions of the pituitary. *Best Pract Res Clin Endocrinol Metab*. 2009;23:639. [PMID: 19945028]

Cohen LE, Radovick S, Wondisford FE. Transcription factors and hypopituitarism. *Trends Endocrinol Metab*. 1999;10:326. [PMID: 10481164]

Crowley RK, Argese N, Tomlinson JW, Stewart PM. Central hypoadrenalism. *J Clin Endocrinol Metab*. 2014;99:4027. [PMID: 25140404]

Faje A. Immunotherapy and hypophysitis: clinical presentation, treatment, and biologic insights. *Pituitary*. 2016;19:82. [PMID: 26186958]

Hardelin JP, Dode C. The complex genetics of Kallmann syndrome: KAL1, FGFR1, FGF8, PROKR2, PROK2, et al. *Sex Dev*. 2008;2:181. [PMID: 18987492]

Hoybye C, Christiansen JS. Growth hormone replacement in adults:current standards and new perspectives. *Best Pract Res Clin Endocrinol Metabol*. 2015;29:115. [PMID: 25617177]

Lauzier F, Turgeon AF, Boutin A, et al. Clinical outcomes, predictors, and prevalence of anterior pituitary disorders following traumatic brain injury: a systematic review. *Crit Care Med*. 2014;42:712. [PMID: 24247474]

Melmed S. Idiopathic adult growth hormone deficiency. *J Clin Endocrinol Metab*. 2013;98:2187-97. [PMID: 23539718]

Molitch ME, Gillam MP. Lymphocytic hypophysitis. *Horm Res*. 2007;68:145. [PMID: 18174733]

Oldfield EH, Merrill MJ. Apoplexy of pituitary adenomas: the perfect storm. *J Neurosurg*. 2015;122:1444. [PMID: 25859802]

Persani L. Clinical review: Central hypothyroidism: pathogenic, diagnostic, and therapeutic challenges. *J Clin Endocrinol Metab*. 2012;97:3068. [PMID: 22851492]

Schneider HJ, Kreitschmann-Andermahr I, Ghigo E, Stalla GK, Agha A. Hypopituitarism. *Lancet*. 2007;369:1461. [PMID: 17467517]

West TA, Sharp S. Neuroendocrine dysfunction following mild TBI: when to screen for it. *J Fam Pract*. 2014;63:11. [PMID: 24475461]

This page intentionally left blank

The Posterior Pituitary (Neurohypophysis)

Alan G. Robinson, MD

ACTH	Adrenocorticotropic hormone		**PAVP**	Plasma vasopressin concentration
ADH	Antidiuretic hormone		**pOsm**	Plasma osmolality
AIDS	Acquired immunodeficiency syndrome		**SIADH**	Syndrome of inappropriate antidiuretic hormone secretion
AVP	Arginine vasopressin			
CSW	Cerebral salt wasting		**V$_{1-3}$**	Vasopressin receptor, types 1-3
DDAVP	Desamino, D-8 arginine vasopressin			

PHYSIOLOGY OF HORMONE FUNCTION

Vasopressin is the water-retaining hormone in all mammals and along with thirst is the primary regulator of osmolality. Pressure and volume are primarily regulated by changes in sodium balance mediated via renin, angiotensin, and aldosterone. The relative importance of vasopressin in regulation of osmolality versus regulation of pressure is reflected in the sensitivity to changes in osmolality versus changes in pressure/volume. Figure 5–1 illustrates the exquisite sensitivity of the osmoreceptors to as little as a 1% change in osmolality. The regulation of vasopressin secretion by baroreceptors, however, involves many concurrent and synergistic sympathetic inputs, and a decrease in volume or pressure of 10% to 15% is necessary before there is a measurable increase in plasma vasopressin.

The exquisite sensitivity in the relation of plasma osmolality to urine osmolality and urine volume is illustrated in Figure 5–2. While the normal range of plasma osmolality encompasses a range of approximately 10 mOsm/L, for any individual, the set-point is much more narrow. Small changes in osmolality produce a corresponding and linear change in plasma vasopressin. As illustrated in Figure 5–2, the normal range of plasma osmolality and plasma vasopressin produces a corresponding linear increase in urine osmolality from maximally dilute to maximally concentrated. This entire range from maximally dilute urine to maximally concentrated urine is accomplished by a narrow range of plasma vasopressin (from approximately 1 to 5 pg/mL). In unusual circumstances, plasma osmolality can rise higher than the normal range, and there

is a corresponding increase in plasma vasopressin, but urine osmolality plateaus at the concentration of the renal inner medulla. When urine in the collecting duct is iso-osmotic with the urine in the inner medulla, maximum urine concentration is achieved. While the relation of plasma osmolality, plasma vasopressin, and urine osmolality are linear, the relation of these to urine volume is not linear. Rather, there is a logarithmic relationship between urine volume and urine osmolality. The total urine volume required to excrete a fixed quantity of urine osmolytes changes relatively little until plasma vasopressin and urine concentration are nearly absent, then urine volume increases dramatically from a few liters per day to 18 to 20 L/d for adults (10 L/m^2).

Also as illustrated in Figure 5–2, thirst has a higher threshold for stimulation than the threshold for vasopressin. This is essential for normal physiology so people can tolerate modest dehydration before becoming thirsty and do not require constant ingestion of fluid.

In the kidney, fluid is conserved by reabsorption of sodium and fluid in the proximal and the distal tubules and then reabsorption of water in the collecting duct. The counter-current multiplier system in the loop of Henle generates a high osmolality in the renal medulla. Vasopressin acts on V$_2$ receptors in principal cells in the collecting duct to stimulate the expression of intracellular water channels, aquaporin-2. When vasopressin binds to the V$_2$ receptor, adenylate cyclase is activated to produce cyclic AMP which stimulates both the production of new aquaporin-2 protein and the transfer of existing aquaporin-2 from the cytoplasm into the cell membrane. In the cell membrane, aquaporins act as water
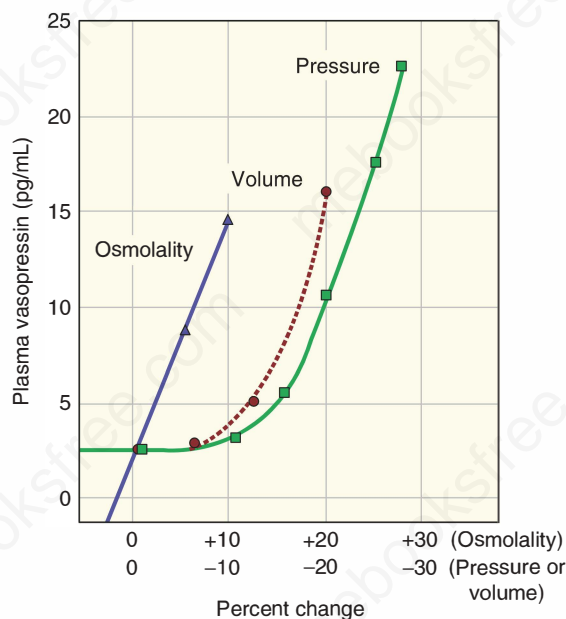
FIGURE 5–1 Vasopressin secretion in response to percentage increases in osmolality or decreases in pressure or volume. (Reproduced with permission from Brenner BM, Rector FC. *The Kidney*. 3rd ed. Philadelphia: WB Saunders; 1986.)

channels in the hydrophobic lipid bilayer so that water moves down the osmotic gradient from the collecting duct into the principal cell. Aquaporins-3 and -4 are involved in moving water out of the principal cell into the medullary extracellular fluid and from there into the circulation. In response to changes in vasopressin, aquaporin-2 can be quickly shuttled in and out of the membrane producing rapid changes in water reabsorption and concentration of urine.

Volume/pressure regulation by vasopressin operates through V_1 receptors on blood vessels. When blood pressure is low and volume is low, activation of V_1 receptors cause contraction of vascular smooth muscle to raise blood pressure and constrict intravascular volume around the available fluid volume.

V_2 receptors also stimulate antihemophilic and von Willebrand factors. There is a third type of receptor (V_3) on anterior pituitary cells that stimulates ACTH. These functions are not further considered in this chapter.

In most physiologic situations, changes in osmolality and volume are additive or synergistic in producing the appropriate physiologic response. For example, most cases of dehydration result in greater loss of water than solute. This produces an increase in plasma osmolality and a decrease in volume that act together to stimulate thirst and the secretion of vasopressin, which promotes retention of water. Similarly, excess ingestion of hypotonic fluid produces a decrease in plasma osmolality and an increase in plasma volume, both of which decrease plasma vasopressin and result in excretion of dilute urine. The normal regulation of osmolality is an elegant and simple system. Fluid ingested and water produced from metabolized food is in excess of true
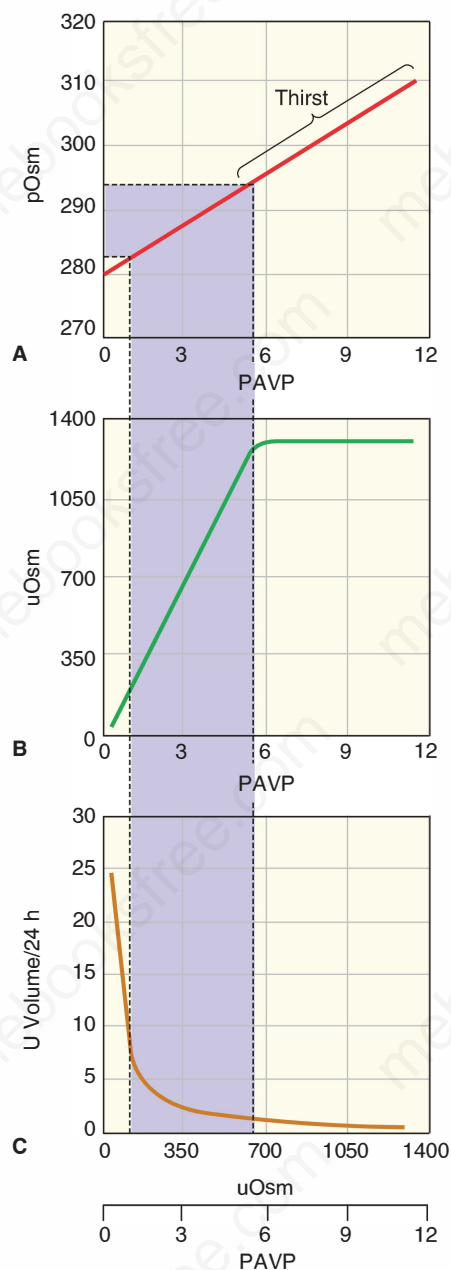
FIGURE 5–2 Normal physiologic relationship between plasma osmolality (pOsm; mOsm/kg H_2O), plasma vasopressin (PAVP; pg/mL), urine osmolality (uOsm; mOsm/kg H_2O), and urine volume (L/d). **(A)** Changes in osmolality induce linear changes in vasopressin with a normal physiologic range of osmolality producing vasopressin levels of 0.5 to 5 pg/mL. **(B)** The physiologic range of vasopressin produces linear changes in urine osmolality. At vasopressin levels above 5 to 6 pg/mL, urine osmolality is at the maximum, determined by the osmolality of the inner medulla of the kidney. **(C)** Assuming a constant osmolar load, the relationship of volume to urine osmolality is logarithmic. The urine volume to excrete a given osmolar load at the urine osmolality in **(B)** is indicated in **C**. (Data from Robertson GL, Shelton RL, Athar S. The osmoregulation of vasopressin. *Kidney Int*. 1976 Jul;10(1):25-37. Copyright 2003, A.G. Robinson, University of California at Los Angeles.)

need. The retained water causes a small decrease in plasma osmolality with a small decrease in vasopressin and excretion of the ingested fluid. If the water intake is not sufficient to supply body needs, plasma osmolality rises, producing concentrated urine to reduce fluid loss, and stimulates thirst that induces drinking to replenish body fluid.

Young infants get all or most of their calories in fluid and they do not dilute their urine as well as adults, so they are at risk to handle excess fluids. Fluid and electrolyte balance is not well regulated in elderly people. By age 80, total body water declines to as low as 50% of the level in normal young adults. Older subjects may have a decrease in glomerular filtration rate, and the collecting duct is less responsive to vasopressin. Numerous studies have indicated a decreased thirst with dehydration in elderly subjects, but also a lessened ability to excrete a water load. These age-related changes in body fluid and renal function predispose elderly people to both hypernatremia and hyponatremia. Elderly people are also more likely to have comorbid diseases that enhance their tendency to become hyponatremic or hypernatremic. It is incumbent upon the physician to pay special attention to fluid balance in children and older people.

ANATOMY OF HORMONE SYNTHESIS AND RELEASE

The posterior pituitary (unlike the anterior pituitary) is not a gland but only the distal axon terminals of the hypothalamic magnocellular neurons that make up the neurohypophysis. The cell bodies of these neurons are located in paired supraoptic nuclei and paired paraventricular nuclei of the hypothalamus. The paraventricular nuclei are located on each side of the third ventricle and axons from these neurons trace laterally and anteriorly to the location of the magnocellular neurons of the supraoptic nucleus just lateral to and above the optic chiasm. Axons of the supraoptic nucleus join axons of the paraventricular nucleus and course to the basal hypothalamus where they join the axons from the other side and course through the infundibular stalk to the axon terminals in the posterior pituitary.

The anatomic location of the physiologic regulators of osmotic and pressure/volume (baroreceptor) are located in vastly different sites. Osmoreceptors that control both thirst and osmotic regulation of vasopressin are located in the hypothalamus just anterior to the third ventricle, so the entire regulation of osmotic-induced changes in thirst and vasopressin secretion resides in a small and discrete area in the hypothalamus. Increases in osmolality stimulate the osmoreceptor to send positive signals to stimulate thirst and to release vasopressin. For volume and pressure regulation, the receptors are located in the chest. There are high-pressure arterial baroreceptors in the carotid sinus and aortic arch and low-pressure volume receptors in the atria and pulmonary venous system. The pressure/volume receptor afferent signals are carried through cranial nerves IX and X and synapse in the medulla before carrying their input to the magnocellular neurons. Both excitatory and inhibitory baroreceptor influences act on the magnocellular neurons, although some evidence indicates that a predominant mechanism involves the tonic inhibition of vasopressin secretion. A decrease in pressure/volume decreases the inhibition and stimulates release of vasopressin.

Vasopressin and oxytocin are nonapeptides that are synthesized in the cell bodies of the magnocellular neurons as part of a larger precursor molecule consisting of the hormone; a specific neurophysin; and (for vasopressin only) a glycopeptide, copeptin. Molecular weights are approximately 1000, 10,000, and 5000 daltons, respectively. Like other polypeptide hormones, the precursor proteins traverse the endoplasmic reticulum and the Golgi apparatus to be packaged in secretory granules. The neurosecretory granules then travel along microtubules down the long axons through the stalk of the infundibulum to the posterior pituitary where the granules are stored. During transport, peptide enzymes (peptidases) within the neurosecretory granules cleave the prohormone into the hormone (vasopressin or oxytocin), the carrier protein (neurophysin), and (for vasopressin) the glycopeptide. The synthesis of oxytocin and vasopressin is in separate neurons organized in clusters within the paraventricular nuclei and supraoptic nuclei. This allows stimulation of hormone specific neurons and independent release of individual hormones. The stimulus for secretion of vasopressin or oxytocin is by neurotransmitters acting on the appropriate magnocellular cell body (perikarya). An action potential propagates along the axon, causing an influx of calcium at the axon terminal that stimulates fusion of neurosecretory granules with the cell membrane and release of the content of the granules into the perivascular space.

PATHOPHYSIOLOGY

All defined pathophysiology of the posterior pituitary is related to the function of vasopressin rather than oxytocin. Both decreased function of vasopressin and increased function of vasopressin manifest as abnormalities of water balance and depend upon action of vasopressin on the V_2 receptor of the kidney rather than the V_1 vascular receptors. The clinical presentation of low vasopressin function (diabetes insipidus) or excess vasopressin function (syndrome of inappropriate antidiuretic hormone secretion [SIADH]) is determined by the physiology of thirst. In humans thirst is well regulated to increase when needed but not well regulated to decrease when not needed. When vasopressin is decreased or absent (diabetes insipidus), there is abnormal excretion of a large volume of dilute urine. This should cause hyperosmolality and an increase in serum sodium. However, as illustrated in Figure 5–2, an elevated plasma osmolality stimulates thirst, and a patient will drink a large volume of fluid to keep their sodium in the high normal range. So, the clinical presentation of decreased vasopressin is one associated with polyuria of dilute urine, together with polydipsia, but with serum sodium in the normal range. Alternatively, when vasopressin is increased, there is abnormal concentration of urine and retention of water. The volume expansion by retained water dilutes serum sodium. If thirst were inhibited by hypo-osmolality as efficiently as thirst is stimulated by hyperosmolality, then the decrease in osmolality would produce a profound decrease in fluid intake, and a normal serum sodium would

be maintained by the natural loss of fluid with perspiration, gastrointestinal fluid loss, and the urine output necessary to excrete ingested osmoles produced by food intake. Thirst, however, is not sufficiently inhibited, and the volume expansion not only dilutes serum sodium, but induces sodium excretion to reduce extracellular volume, which worsens the hyponatremia. So, the presentation of increased vasopressin is biochemically a presentation of hyponatremia. There is concentrated urine with no change in thirst, modest volume expansion, hyponatremia, and natriuresis.

The other physiologic principle that determines some of the pathophysiology of abnormal secretion of vasopressin is that while osmolality is regulated by changes in water balance, and volume is regulated by changes in sodium balance; when in conflict, volume will be preserved at the expense of osmolality. Thus, if dehydration is produced by diabetes insipidus, sodium will be retained to protect volume, even if it aggravates hypernatremia. With SIADH when sodium is diluted and extracellular volume is expanded, natriuresis is initiated to decrease extracellular volume in spite of the need to retain sodium to correct the hypo-osmolality.

DEFICIENT VASOPRESSIN: DIABETES INSIPIDUS

Diabetes insipidus is literally the excretion of a large volume of urine (diabetes) that is hypotonic, dilute, and tasteless (insipid). Patients present with polyuria. Other causes of polyuria such as osmotic diuresis that occur in diabetes mellitus or intrinsic renal disease must be excluded. Frequent urination, without an increase in urine volume, suggests a urologic abnormality. Most adults will tolerate polyuria without complaint until it exceeds 3 to 4 L/d. If polyuria is shown to be dilute, the pathophysiologic mechanisms include: (1) primary ingestion of excess fluid, (2) abnormally decreased synthesis and secretion of vasopressin, (3) increased metabolism of vasopressin, and (4) decreased end-organ response to vasopressin.

A. Ingestion of excess fluid and normal urine excretion—primary polydipsia
Primary polydipsia with ingestion of large quantities of water produces modest decreases in plasma osmolality, decreased secretion of vasopressin, and excretion of profound quantities of urine. As the volume of dilute fluid delivered to the collecting duct in adults is approximately 18 L/d (10 L/m^2), this quantity of dilute urine can be excreted on a daily basis while maintaining serum osmolality in the normal range. Virtually any pathologic process in the hypothalamus that causes diabetes insipidus may rarely cause primary stimulation of the thirst center and primary polydipsia on an organic basis. Alternatively, primary polydipsia may occur as a behavioral abnormality in psychiatric patients in whom the excess water ingestion and urination may represent delusions tied to bodily cleansing. The large volume of urine passing through the collecting duct in these patients washes urea out of the inner medulla and reduces the inner medulla osmolality. Additionally, the chronic suppression of vasopressin release and lack of action of vasopressin on the renal collecting duct decreases uptake of urea and decreases the synthesis of aquaporins in the collecting duct cells, adding to the inability to concentrate the urine maximally. Each of these abnormalities returns to normal in several days to weeks after fluid ingestion is decreased to normal.

Primary polyuria may be induced in a normal person if excess fluid is administered parenterally. This occasionally occurs in postoperative pituitary surgery and even in cases that do not involve the pituitary. In many surgical procedures, the stress of the surgery is a stimulus to release of vasopressin, and administered fluids may be retained during the procedure. Postoperatively, the stress is released, and patients have a normal diuresis of the retained fluids. If this normal excretion of fluid is matched by increased fluids administered parenterally, the patient will continue to excrete large volumes of urine. If the administered fluid is normal saline, the patients will excrete a large volume of isotonic urine. If the volume administered and subsequent urine output is sufficiently great, the patients will be unable to concentrate or dilute their urine above or below an iso-osmotic level regardless of therapeutic agents. The large osmotic load delivered to the kidney may make the patient unresponsive even to administered desmopressin, producing an iatrogenic therapeutic dilemma.

B. Abnormal synthesis and secretion of vasopressin
It is important to remember that the complaint of polyuria is based on the volume of the urine not the urinary concentration. Therefore, as evident from Figure 5–2, considerable loss of ability to secrete vasopressin can occur before there is much loss of ability to concentrate the urine and even greater loss can occur before there is a noticeable increase in urine volume.

1. GENETIC ABNORMALITIES OF NEUROHYPOPHYSEAL NEURONS

Familial hypothalamic diabetes insipidus.   Familial diabetes insipidus may be caused by an autosomal dominant mutation in the vasopressin gene. Usually the mutation involves DNA sequences in the neurophysin or signal peptide region of the precursor gene rather than the region encoding vasopressin itself. These mutations cause abnormal folding of the precursor protein, which produces abnormal trafficking and accumulation of a mutant prohormone in the endoplasmic reticulum. This produces alterations in packaging of the prohormone into neurosecretory granules in the Golgi apparatus. By ill-defined mechanisms, this leads to cell death of the vasopressin-producing neurons. Because the pathology in these neurons develops over time, young children may have normal urine output, and diabetes insipidus is not expressed until late childhood. An autosomal dominant form of diabetes insipidus also occurs in association with diabetes mellitus, optic atrophy, and deafness (DIDMOAD), Wolfram syndrome. Children with midbrain malformations may have abnormal structural architecture in the hypothalamus producing diabetes insipidus and abnormalities of thirst.

2. PATHOLOGIC LESIONS OF THE NEUROHYPOPHYSIS

Solid tumors and hematologic malignancies.   The most common solid tumor to produce diabetes insipidus is craniopharyngioma. Suprasellar germinoma or pinealoma in children, which may have elevated human chorionic gonadotropin (hCG) levels and be associated with precocious puberty, commonly produces diabetes insipidus. Metastatic disease to the pituitary hypothalamic area, for example, from breast or lung cancer, is more likely to produce diabetes insipidus than a deficiency of anterior pituitary hormones because the metastases lodge in the portal system of the hypothalamus, where the vasopressin axons from the two sides join to form the pituitary stalk. Usually, there are widespread metastatic lesions elsewhere. Lymphoma or infiltration of the hypothalamus with leukemia are rare causes of diabetes insipidus.

Trauma or surgery.   The anatomy of the neurohypophyseal system aids our understanding of this pathology. Hormone synthesis and regulation of release are high in the hypothalamus. Surgery or trauma at the level of the pituitary traumatizes only the axon terminals. Trauma of the axon terminals may disrupt the release of hormone, but usually this produces only a transient inability to secrete hormone and transient diabetes insipidus. Surgery in the hypothalamic area is more likely to produce abnormal vasopressin secretion. Sectioning of the posterior pituitary stalk produces a characteristic pattern of diabetes insipidus known as the triphasic response. This is illustrated in Figure 5–3. Initial axonal shock inhibits the release of any vasopressin, and there is a period of diabetes insipidus that lasts for 5 to 10 days. Then, the severed axons in the posterior pituitary

become necrotic, and there is uncontrolled release of vasopressin. If excess fluid is administered during this time, the syndrome of inappropriate antidiuretic hormone is produced with hyponatremia (described later). This lasts for 5 to 10 days until all of the residual vasopressin leaks from the axon terminals, and there is then return of diabetes insipidus. While the classic response to stalk section is the three phases, in many clinical situations not all three phases occur. The most common occurrence after surgery in the pituitary or hypothalamus is transient diabetes insipidus that lasts a few days without subsequent sequelae. With partial damage to the pituitary stalk rather than total section, it is possible that the second phase will occur without preceding or subsequent diabetes insipidus (ie, isolated second phase). In the isolated second phase, there are sufficient normally functioning vasopressinergic neurons to avoid diabetes insipidus in the first and third phase; however, the damaged axon terminals leak vasopressin that will, when coupled with excess fluid intake, produce SIADH and hyponatremia. Characteristically, hyponatremia occurs 5 to 10 days after pituitary surgery, lasts for 5 to 10 days, and then resolves. Late occurrence of diabetes insipidus, that is, third phase without a first or second phase, is distinctly uncommon.

Closed head trauma, especially with fracture of the basal skull, may produce diabetes insipidus and the patterns above. Hypothyroidism and adrenal insufficiency both decrease water excretion and when present may mask diabetes insipidus. Normal thyroid and adrenal function must be assured in considering operative or trauma-induced diabetes insipidus, and loss of these hormones should be suspected in the decreasing urine volume interpreted as "recovery" of diabetes insipidus in this setting.

Patients with diabetes insipidus due to surgery or trauma may eventually recover. The level of section determines the number of magnocellular neurons that are actually destroyed. The higher in the stalk and the closer the section to the perikarya, the greater the number of neurons that die. Vasopressin-producing neurons whose axons terminate in the hypothalamus and serve as secretagogues for ACTH in the anterior pituitary may persist and hypertrophy after surgery or trauma. Additionally, branching of axons may develop in the hypothalamus after surgery or trauma, and these branches may generate new connections of vasopressinergic neurons to blood vessels. By these mechanisms, sufficient vasopressin function may return (usually within 1 year) to have normal fluid balance and no symptoms of diabetes insipidus.

Granulomatous and infectious disease.   Langerhans cell histiocytosis is the generic term that includes severe fulminant visceral Letterer-Siwe disease, multifocal Hand-Schuller-Christian disease, and benign eosinophilic granuloma. Diabetes insipidus is part of the central nervous system involvement in these disorders and is associated with other abnormalities of the head involving the cranial bones, oral mucosa, or brain. Systemic manifestations in the lung, bone, and skin may also be present. Wegener granulomatosis and sarcoidosis may also have diabetes insipidus as a part of the central nervous system pathology. Tuberculosis (historically) and other infections at the base of the brain may cause diabetes insipidus.

FIGURE 5–3 Illustration of the phases of urine output after section of the pituitary stalk. The *triphasic response* consists of: (1) diabetes insipidus due to axonal shock and lack of release of vasopressin; (2) an antidiuretic interphase when vasopressin leaks from the severed neurons; and (3) return of diabetes insipidus when the store of vasopressin in the posterior pituitary is depleted.

Lymphocytic infundibulohypophysitis.   It is now recognized that many cases of diabetes insipidus, that were previously termed idiopathic and for which there was no specific etiology, are probably due to lymphocytic infiltration of the neurohypophysis on an autoimmune basis. Magnetic resonance or computed tomographic imaging may reveal enlargement of the pituitary stalk as evidence of the lymphocytic invasion. There may be resolution of the autoimmune response with time, and the stalk may return to normal size, but the diabetes insipidus is usually permanent.

Essential hypernatremia—Adipsic hypothalamic diabetes insipidus.   A rare variant of diabetes insipidus involves an absent osmoreceptor function, but an intact baroreceptor function. In this form of diabetes insipidus, if patients drink they have polyuria. However, the patients do not sense thirst and, therefore, usually do not drink. Lack of fluid intake results in an elevation of serum osmolality, but osmoreceptor function is absent and does not respond to the increased osmolality by secreting vasopressin or stimulating thirst. Vasopressin synthesis and storage are normal as demonstrated by specific tests of baroreceptor function. Characteristically, these patients do not drink water and hence become sufficiently dehydrated that their baroreceptors stimulate vasopressin release; they remain in a balanced situation with an elevated serum sodium, relatively concentrated urine, and lack of thirst.

Cerebral anoxia/brain death.   Brain dead patients who are on life support often have diabetes insipidus as part of the central nervous system pathology. It is standard practice to maintain electrolyte and water balance for maximal preservation of organs for transplantation.

C.   Increased metabolism of vasopressin in pregnancy to produce diabetes insipidus   During pregnancy, there is a true resetting of the osmostat at an osmolality approximately 10 mOsm/kg H_2O less than the normal pOsm–PAVP relationship. So, in Figure 5–2 (A), the pOsm–PAVP would have the same linear relationship but would be shifted downward on the graph with the increase in PAVP beginning at pOsm of 270. Both increases and decreases in plasma vasopressin occur at the lower pOsm while the uOsm and Urine volume response to PAVP remains the same. There is a similar change in the regulation of plasma volume. In normal pregnancy, total body water is increased by 7 to 8 L as a result of profound vasodilatation. This expanded volume is also sensed as normal with vasopressin increasing and decreasing around this new volume. Chorionic gonadotropin (hCG) and relaxin, which are produced in pregnancy, are the mediators of these effects. In pregnancy the placenta produces an enzyme, cysteine aminopeptidase, that is known as oxytocinase but similarly degrades vasopressin. Therefore, vasopressin metabolism is increased in pregnancy, markedly so from 20 to 40 weeks of gestation.

Because of the reset osmostat, a "normal" serum sodium during pregnancy may represent dehydration. There are two types of diabetes insipidus that are associated with pregnancy. In the first, the patient has a preexisting limitation of vasopressin function such as partial hypothalamic diabetes insipidus or mild nephrogenic diabetes insipidus. The ability to concentrate the urine may be limited but sufficient to maintain acceptable urine volume prepartum. During pregnancy the metabolism of vasopressin is accelerated. With increased metabolism of vasopressin, these patients may then not have sufficient concentrating ability to maintain an acceptable urine volume (see vasopressin–volume relationships in Figure 5–2) and manifest diabetes insipidus. When the pregnancy ends, PAVP is again sufficient so that the patient reverts to the existing asymptomatic condition. A second form of diabetes insipidus occurs when the level of cysteine aminopeptidase is extraordinarily higher than in normal pregnancy and produces diabetes insipidus in a patient with normal pituitary and renal function. There may be other associated pathology with the latter disease such as preeclampsia, fatty liver, and coagulopathies. Diabetes insipidus in this situation also abates when the pregnancy ends and may not recur with subsequent pregnancies.

D.   Decreased end organ response to vasopressin—nephrogenic diabetes insipidus

1. **Congenital nephrogenic diabetes insipidus**   There are two causes of congenital nephrogenic diabetes insipidus: (1) an X-linked recessive mutation of the V_2 receptor, which accounts for 90% of the cases; and, (2) an autosomal recessive mutation of the aquaporin-2 water channels. The X-linked disorder, of course, only occurs in males while disease due to mutations of the aquaporin-2 gene occurs in males and females. Female carriers of the X-linked V_2 receptor abnormality usually have no clinical disease nor do heterozygote carriers of the recessive aquaporin-2 mutations. Nephrogenic diabetes insipidus due to aquaporin-2 defects is produced when the child inherits a mutated gene from each parent that carries the recessive mutation.

    Regardless of etiology, the presentation of nephrogenic diabetes insipidus usually occurs in the first week of life presenting with vomiting, constipation, failure to thrive, fever, and polyuria. The patients are found to be hypernatremic with low urine osmolality. When measured, vasopressin levels are elevated. Immediate treatment is essential for normal neurologic development.

2. **Acquired nephrogenic diabetes insipidus**   The ability to produce concentrated urine depends upon hyperosmolality of the inner medulla which in turn requires normal kidney architecture with descending and ascending limbs of the loop of Henle, normal sodium transport, functional aquaporins, and intact long loop vascular structures (vasa recta) so that the hyperosmolality in the inner medulla is not washed away by normal blood flow. One form of nephrogenic diabetes insipidus is produced by renal disease that distorts the vascular architecture of the kidney and washes out the inner medulla; for example, polycystic renal disease, infarcts, sickle cell anemia, and so forth. The polyuria associated with potassium deficiency and hypercalcemia are both associated with decreased expression of aquaporin-2. Among drugs that may produce nephrotoxicity, the most common to produce nephrogenic diabetes insipidus are lithium (commonly when used to treat bipolar affective disorders) and demeclocycline.

DIAGNOSTIC TESTS OF DIABETES INSIPIDUS

One can determine whether there is true polyuria by measuring the 24-hour urine output, but collecting this large volume is inconvenient. Alternatively, the patient can keep a diary, recording the time and volume of each voided urine without saving the urine for examination.

A. Dehydration test   After other causes of polyuria are eliminated, the diagnosis of diabetes insipidus is accomplished by some form of dehydration test. If the patient is dehydrated when first seen and the urine osmolality is low, the "test" has been done. Plasma vasopressin is measured but the specific testing described later is not necessary. When indicated, a dehydration test should be done in a controlled environment. The description here is for adults. Dehydration of children requires special attention and should only be done by a pediatrician. Dehydration tests should not be done in infants. The patient is weighed at the beginning of the test, and a serum sodium concentration and plasma osmolality are obtained. Nothing is allowed by mouth, and the volume and osmolality of each voided urine is recorded. The patient is weighed after each liter of urine is excreted. When two consecutive measures of urine osmolality differ by less than 10% and the patient has lost 2% of his body weight, the serum sodium concentration and plasma osmolality are again determined, and a blood sample is drawn for a determination of plasma vasopressin. In the original description the patient was then given 5 units of aqueous vasopressin parenterally, although now it is often 2 μg of desmopressin, a synthetic analogue of vasopressin described later. Urine osmolality and output are recorded for an additional 2 hours. (Note that pediatricians do not use desmopressin and use a lower dose of vasopressin in this test to avoid hyponatremia.) The duration of the test varies, but in most patients a plateau in the urine osmolality will be reached by 18 hours. The test is discontinued if the patient loses 3% of the body weight. The different responses to this testing in adults are depicted in Figure 5–4. Blood is drawn to measure plasma vasopressin, and the response to administered vasopressin or desmopressin is determined.

Patients with primary polydipsia are distinguished from diabetes insipidus by the concentration of the urine at a plateau which is frequently 500 to 700 mOsm/kg, which is less than normal maximum, and by the lack of further response to administered desmopressin. Patients with hypothalamic diabetes insipidus have undetectable or low plasma vasopressin levels at the end of dehydration, minimal concentration of the urine, and a significant increase in urine osmolality to administered desmopressin, at least 50% but often 200% to 400%. Patients with nephrogenic diabetes insipidus do not concentrate their urine similar to hypothalamic diabetes insipidus, but the levels of plasma vasopressin are high, often greater than 5 pg/mL at the end of the test. Additionally, there is no further increase in urine osmolality after the administration of desmopressin. Occasionally, patients with partial hypothalamic diabetes insipidus concentrate their urine minimally with dehydration, but their maximum urinary concentration
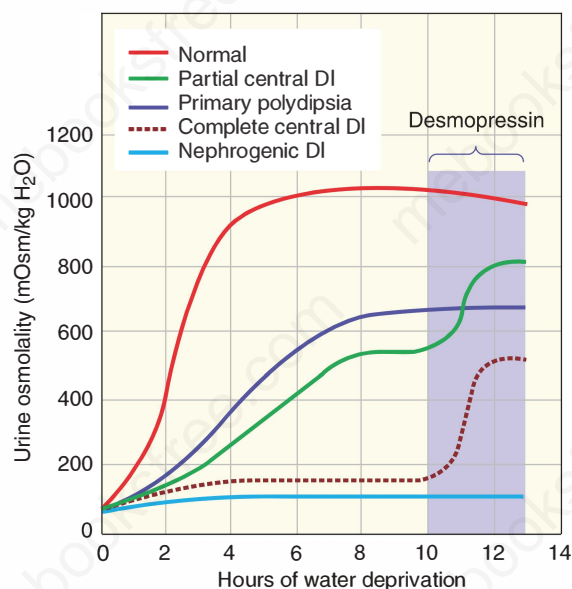
FIGURE 5–4   Urine osmolality during a dehydration test followed by administration of desmopressin to differentiate various types of diabetes insipidus and primary polydipsia as described by Miller et al. (Miller M, Dalakos T, Moses AM, Fellerman H, Streeten D. Recognition of partial defects in antidiuretic hormone secretion. *Ann Int Med*. 1970;73:721.) See text for discussion of the differential diagnosis based on the plateau of urine osmolality with dehydration and subsequent response to administered desmopressin (or vasopressin, see text).

is not achieved, and there is an additional boost with administered desmopressin. This pattern can be similar to some patients with primary polydipsia who achieve a plateau level of urine osmolality before reaching their maximum attainable urine osmolality and show a boost in urine osmolality in response to desmopressin. These patients are usually distinguished by their response to desmopressin over time, which should be considered a continuation of the diagnostic test. Patients with partial hypothalamic diabetes insipidus decrease their urine volume and their thirst while those with primary polydipsia may continue to drink excessive amounts of water and develop hyponatremia.

Measurement of plasma vasopressin by clinically available assays is only of help in distinguishing nephrogenic diabetes insipidus (in which the level will be unequivocally elevated) from other causes of diabetes insipidus. Recent reports have described a radioimmunoassay of copeptin, the inactive carboxy peptide that is cleaved from the provasopressin molecule and secreted with vasopressin and neurophysin. The clinical usefulness of this measure is under investigation.

B. Imaging in diabetes insipidus   Imaging of the hypothalamus is an important diagnostic tool in diabetes insipidus. The exquisite sensitivity of the kidney to vasopressin, especially with regard to urine volume, determines that even 10% of the vasopressin neurons may be sufficient to maintain homeostasis without symptoms. Because of the wide anatomic distance between the paraventricular and supraoptic nuclei, tumors that

produce diabetes insipidus are either sufficiently large to destroy 90% of the vasopressin neurons or, most commonly, are located just above the diaphragma sella where the neurohypophyseal tracts of the two sides converge as they enter the pituitary stalk. Small tumors of the anterior pituitary confined to the sella are never the cause of diabetes insipidus. On T1-weighted images, the normal posterior pituitary appears as a bright spot on MRI studies. The bright spot is produced by the stored hormone in the posterior pituitary. Most normal subjects have the bright spot, although it decreases with age, and most patients with diabetes insipidus do not have a bright spot. An MRI may reveal thickening of the stalk and absence of the posterior pituitary bright spot. This is a characteristic finding of lymphocytic infundibulitis and of some granulomatous diseases. The latter prompts a search for granulomatous disease elsewhere in the body.

TREATMENT OF DIABETES INSIPIDUS

As noted earlier, adults with intact thirst mechanisms usually drink sufficient fluid to maintain serum sodium levels in the normal range. Babies and adults without access to fluid or patients who are unconscious may not drink and can develop severe dehydration and hypernatremia. Many pathologic entities that damage the hypothalamus and cause diabetes insipidus, for example, lymphocytic hypophysitis, granulomatous disease, metastatic disease, will be treated, but rarely does successful treatment of the underlying disease produce remission of the diabetes insipidus.

Hypothalamic diabetes insipdus is best treated by the vasopressin analogue desmopressin (desamino, D-8 arginine vasopressin, DDAVP). In desmopressin, the amine at position 1 is removed, increasing the half-life of the compound, and at position 8, L-arginine is changed to D-arginine, markedly reducing the pressor activity. Thus, the agent is highly selective for the V_2 receptor and based on dose and duration produces nearly 2000 times more antidiuresis than does the naturally occurring L-arginine vasopressin. Desmopressin is available as tablets (0.1 and 0.2 mg), a solution for nasal installation (100 μg/mL), a parenteral solution (4 μg/mL) and in Europe DDAVP Melt, an oral lyophilisate (60, 120, and 240 μg). There is considerable individual variability in the duration of action of desmopressin, so the dosing should be adjusted for each individual. Usually, a satisfactory schedule can be achieved with 2 to 3 tablets per day or one to two intranasal doses. A parenteral preparation can be given intravenously, intramuscularly, or subcutaneously and is 5 to 20 times more potent than the intranasally administered dose. The drug is expensive, and in many patients, a smaller dose given more frequently is more cost-effective. Other agents that prolong the action of vasopressin or the release of vasopressin include chlorpropamide, carbamazepine, clofibrate, and indomethacin, but these are off-label uses and not first-line therapies. Desmopressin is the only therapeutic agent recommended for treatment of diabetes insipidus in pregnancy because desmopressin is not destroyed by cysteine aminopeptidase and has minimal oxytocic activity on the uterus.

In patients with inadequate thirst, the best treatment is continuous administration of desmopressin to maintain a concentrated urine and then rigid adjustment of the amount of fluid that must be taken every 6 to 8 hours to maintain normal serum sodium.

In nephrogenic diabetes insipidus, any offending drug or electrolyte abnormality that might produce acquired nephrogenic diabetes insipidus should be stopped or corrected. In congenital nephrogenic diabetes insipidus, therapy is aimed at reducing urine volume through a low sodium diet and a thiazide diuretic. This causes a natriuresis which produces some contraction of the extracellular fluid volume, decreased glomerular filtration rate, decreased delivery of fluid to the collecting duct, and a decreased urine volume. Amiloride is especially recommended in this setting because it is potassium-sparing. Amiloride may also have some advantage in lithium-induced nephrogenic diabetes insipidus because amiloride decreases lithium entrance into cells in the distal tubule. Indomethacin has an antidiuretic action that especially prolongs the action of vasopressin and administered desmopressin. It also decreases urine volume in nephrogenic diabetes insipidus, but there is concern about gastrointestinal bleeding.

When diabetes insipidus occurs in patients who also have anterior pituitary deficiency, adequate treatment with thyroid hormone and hydrocortisone is essential to maintain normal renal response to desmopressin. Clinical situations such as surgical procedures, treatments that require a saline diuresis, and periods when patients are not allowed fluids by mouth require careful balance of antidiuretics (often a low dose of vasopressin by infusion), administered fluid, and sodium.

EXCESS VASOPRESSIN: SYNDROME OF INAPPROPRIATE ANTIDIURETIC HORMONE

Excess vasopressin becomes a clinical problem when there is concurrent retention of water producing hypo-osmolality. Hypo-osmolality is common in hospitalized patients but in most cases is due to underlying illness. Statistically, hyponatremia is associated with increased morbidity and mortality because it is a measure of the severity of the underlying illness. This chapter focuses on cases in which inappropriately elevated vasopressin and water retention are the cause of the hypo-osmolality (see also Chapter 21: Humoral Manifestations of Malignancy).

The differential diagnosis of hypo-osmolality is based on measuring the serum sodium concentration and estimating the extracellular fluid status. Sodium is the major cation of the extracellular fluid, and potassium is the major cation of the intracellular fluid. Water moves freely between the intracellular and extracellular fluid, so osmolality is always equivalent in both compartments. The relative balance between solute depletion and water retention as a cause of hypo-osmolality is complex, but in virtually all cases of clinically significant hypo-osmolality, there is some element of impaired water excretion and/or excessive water intake.

Plasma osmolality can be measured directly by freezing point depression or vapor pressure of plasma, but usually, hypo-osmolality is brought to clinical attention because of decreased

concentration of sodium rather than measure of osmolality. In the extracellular fluid, the major contributions to osmolality are sodium, glucose, and blood urea nitrogen (BUN). A calculated osmolality usually correlates well with measured osmolality.

$$\text{Plasma osmolality (mOsm/kg H}_2\text{O)}$$
$$= 2 \times [\text{Na}^+] \text{ (mEq/L)} + \text{Glucose (mg/dL)}/18$$
$$+ \text{BUN (mg/dL)}/2.8$$

Occasionally, when the serum sodium concentration is measured by flame photometry the measured sodium is artifactually low because flame photometry calculates the sodium in a fixed volume of plasma. If a large proportion of the plasma volume is taken up by extremely elevated levels of lipid or protein, sodium determined by flame photometry is low. Plasma osmolality determined by freezing point or vapor pressure is a direct measure of particles in solution and will be normal in these situations. So the low level of sodium by flame photometry is referred to as pseudo-hyponatremia. Hyperglycemia will produce hyponatremia because of the shift of water from the intracellular fluid to the extracellular fluid; however, the calculated osmolality will be normal. The serum sodium can be corrected for the glucose elevation by adding 1.6 to 2.4 mEq/L for each 100 mg/dL increase in serum glucose.

When true hypo-osmolality is found to exist, the differential diagnosis is of hyponatremia as illustrated in Table 5–1. The disorder is divided into four major subgroups based on the extracellular fluid volume status and the measured urinary sodium. Extracellular fluid volume status may be determined by central venous pressure, BUN/creatinine, hematocrit, or plasma protein concentrations, but usually is estimated by physical examination. If the patient is dehydrated and the urinary sodium is low, this indicates normal physiologic response to extra-renal sodium loss such as vomiting or diarrhea with continued intake of water. This is the most common cause of hyponatremia in children. The appropriate therapy is to replace the sodium and fluid deficiency with normal saline. If the patient is dehydrated but the urinary sodium is increased, this indicates a renal loss of sodium inappropriate to the decreased volume and hyponatremia. This may be due to intrinsic renal disease, diuretic use, aldosterone deficiency

TABLE 5–1 Differential diagnosis of hyponatremia.

Volume	Urine Na$^+$	Diagnosis
Low (evidence of dehydration)	<20 mEq/L	Category 1: Total body Na$^+$ depleted, normal renal response (eg, hemorrhage, GI losses)
	>25 mEq/L	Category 2: Renal Na$^+$ loss (eg, renal disease, diuretics, CSW Addison disease)
Normal or expanded (edema may be present)	<20 mEq/L	Category 3: Hyperaldos teronism secondary to inadequate perfusion (eg, CHF ascites, etc)
	>40 mEq/L	Category 4: SIADH-Na$^+$ loss secondary to volume expansion

Copyright 2005, A.G. Robinson, University of California at Los Angeles.

(see discussion of Addison disease in Chapter 9), or cerebral salt wasting (CSW) (to be discussed later). Appropriate therapy is to replace the sodium and fluid loss with normal saline but also with appropriate treatment of the underlying defect. If the extracellular fluid volume is expanded with edema or ascites and the urinary sodium is low, this indicates hyperaldosteronism, secondary to reduced or ineffective plasma volume as in cirrhosis or congestive heart failure, and the appropriate therapy is treatment of the underlying condition. Where the extracellular fluid volume appears to be normal with increased urinary sodium, this indicates the pathophysiology of inappropriate secretion of antidiuretic hormone. Note that the urine osmolality will not be maximally dilute and that measured vasopressin will not be maximally suppressed in any of the four categories in Table 5–1. Although not appropriate to the osmolality, in the first three categories, the elevation of vasopressin is appropriate to the real or perceived decreased plasma volume. It is only in the fourth category (SIADH) that vasopressin is inappropriate both for plasma volume and osmolality. Copeptin, the inactive glycopeptide that is cleaved from the provasopressin molecule and secreted with vasopressin and neurophysin, was mentioned earlier. Some studies have indicated that measurement of levels of this peptide might be useful in evaluating various causes of hyponatremia.

A. The pathophysiology of the hyponatremia in SIADH   The pathophysiology of SIADH begins with uncontrolled secretion of vasopressin. Thirst is not adequately suppressed so fluid intake continues. This produces volume expansion and hyponatremia in the sequence illustrated in Figure 5–5. Vasopressin limits the excretion of water in the urine. Continued ingestion of water (**part A**) produces an expansion of extracellular and intracellular volume. The body attempts to bring the extracellular fluid volume back to normal by natriuresis (**part B**) of isotonic urine. The mechanism of natriuresis is complex and involves increased glomerular filtration; pressure natriuresis; and natriuretic factors, especially atrial natriuretic peptide and brain natriuretic peptide. This natriuresis decreases total body water and total body sodium but because it is isotonic, it contributes little to the degree of hyponatremia. Next, the body attempts to return intracellular fluid volume to normal by (**part C**) excreting from the intracellular fluid potassium and organic osmolytes such as glutamine, glutamate, myoinositol, aspartate, and N-acetylaspartate. In spite of the attempt to normalize extracellular and intracellular fluid, there remains a tendency for these compartments to be slightly expanded. The last adaptation (**part D**) is caused by this tendency for volume expansion and produces changes in the kidney to make it less responsive to the chronic inappropriate excess of vasopressin and to allow an increase in water excretion. Vasopressin retains water by stimulating V$_2$ receptors on the principal cells of the collecting duct. This stimulation both increases the synthesis of aquaporin-2 molecules and the insertion of aquaporins into the cell membrane. With chronic excess of vasopressin, the density of aquaporins in the membrane increases dramatically, producing a state of chronic increase in water retention. However, in response to the chronic volume expansion (produced by the water

FIGURE 5–5 Illustration of the changes in intracellular fluid and extracellular fluid volumes in the pathophysiology of SIADH with the chronologic pattern of: **(A)** Expanded ECF and ICF volume with solute dilution. **(B)** Natriuresis to decrease ECF volume expansion. **(C)** Extrusion of intracellular potassium and osmolytes to decrease intracellular volume expansion. **(D)** Restoration of near normal volume in ECF and ICF. **(E)** Renal adaptation to allow excretion of more dilute urine in spite of excess vasopressin. (Copyright 2005, A.G. Robinson, University of California at Los Angeles.)

retention) there is by unknown mechanisms an adaptation that reduces the number of aquaporins in the principal cells. In this new state **(part E)**, ingestion of sodium will somewhat re-expand the extracellular fluid volume and be excreted, while ingestion of water will be more easily excreted because of the renal adaptation of decreased aquaporins—therefore a new steady state.

Of the diagnoses of hyponatremia shown in Table 5–1, the two categories that are most difficult to differentiate are those with elevated urinary sodium concentrations. On physical examination, it may be difficult to differentiate moderately low extracellular fluid volume from normal extracellular fluid volume. In this situation, the differential diagnosis may be aided by a volume challenge of normal saline infused at a modest rate over a few hours while following urinary and plasma sodium determinations. If the patient is in diagnostic category 4 (SIADH) and relatively equilibrated, the administered saline will be excreted so there will be an increase in urinary sodium and no change or a slight decrease in serum sodium concentration. If the patient is in category 2 with a renal loss of sodium, sodium from the administered

saline will be retained, and the water excreted will somewhat dilute urinary sodium. This will result in a decrease in urinary sodium while the serum sodium concentration will rise. This volume challenge in difficult cases is not considered therapy, but only an assist to make the appropriate diagnosis, which will then lead to the initiation of appropriate therapy.

In diagnostic category 2 (Table 5–1), there is considerable controversy regarding the diagnosis and incidence of CSW. CSW is severe primary natriuresis producing volume depletion and hyponatremia. The diagnosis is only confirmed when volume depletion can be confirmed by some combination of decreased central venous pressure, decreased plasma volume, increased BUN/creatinine, or increased hematocrit and plasma protein levels. CSW has been intensively studied as a cause of hyponatremia in subarachnoid hemorrhage. Natriuretic peptides are increased in subarachnoid hemorrhage, but they are also increased in well-documented SIADH. Measured plasma volume in subarachnoid hemorrhage is difficult to interpret because of the lack of sensitivity of readily available clinical measures. There may be both inappropriate

secretion of vasopressin and inappropriate natriuresis in these patients with the clinical presentation depending on the relative contributions of each. It is consistent that infusion of isotonic saline does not correct the hyponatremia in SIADH but may do so with CSW. The most convincing cases of CSW are those that occur in hospitalized patients in whom there is a sudden unexplained increase in urine volume (not seen in SIADH) and urine sodium excretion. This has been reported in several cases of children with traumatic brain injury or brain surgery. The acute massive diuresis and natriuresis is accompanied by definitive volume contraction by the measures listed earlier. Most of these cases respond to replacement with normal or hypertonic saline.

B. Symptoms of hyponatremia   The symptoms of hyponatremia are largely dependent on the rapidity of the development of hyponatremia. When hyponatremia develops rapidly and is severe (serum sodium level <120 mEq/L), patients are at risk for cerebral edema with herniation of the brain stem (especially children and young women). Other complications include neurogenic pulmonary edema, seizures, coma, and respiratory arrest. Hyponatremia that develops slowly over a long period of time is surprisingly well-tolerated even at very low levels of serum sodium. Neurologic symptoms usually do not occur with sodium values above 120 mEq/L, but any degree of hyponatremia might exacerbate other comorbid conditions. Hyponatremia is usually considered chronic if hyponatremia has developed slowly and persisted for greater than 48 hours.

C. Clinical syndromes of SIADH   The clinical criteria for SIADH remain those described by Bartter and Schwartz in 1967: decreased plasma osmolality; inappropriate concentration of the urine; clinical euvolemia; increased urinary sodium excretion; and absence of other causes of euvolemic hypo-osmolality such as hypothyroidism, adrenal insufficiency, or diuretic use. The three main etiologies of SIADH are ectopic production of vasopressin by cancer (see Chapter 21), drug-induced SIADH, and lesions in the pathway of the baroreceptor system, especially central nervous system disorders and pulmonary disorders. As noted earlier, the baroreceptors consist of a diffuse system of receptors in the chest and synapses in the brain. Some of the input from them is inhibitory. Lesions which disrupt the flow of signals in the lung or in the brain may decrease this inhibitory signal and produce inappropriate secretion of vasopressin. Acquired immune deficiency syndrome (AIDS) is a recognized cause of SIADH, although dehydration, adrenal insufficiency, pneumonitis, and renal tubular toxicity associated with therapy may confound the ability to make the diagnosis of SIADH an exclusive cause of the hyponatremia. Prolonged strenuous exercise (eg, marathon running) may increase secretion of vasopressin and with excessive intake of hypotonic fluid induce SIADH. A rare cause of SIADH reported in two pediatric patients is a genetic mutation of the V_2 receptor in principal cells producing unregulated activation. These patients have the manifestations of hyponatremia and SIADH but no measurable plasma vasopressin. Some elderly patients fit the criteria for SIADH without a definitive diagnosis.

TABLE 5–2   **Major etiologies of SIADH.**

Ectopic production of vasopressin by tumors
Bronchogenic carcinoma
Carcinoma of duodenum and pancreas
Ureteral, prostate, bladder carcinoma
Carcinoma of the uterus
Thymoma
Mesothelioma
Lymphoma and leukemia
Drug induced
Desmopressin and oxytocin
Chlorpropamide
Carbamazepine
Clofibrate
Ecstasy
Selective serotonin reuptake inhibitors
Tricyclic antidepressants
Monoamine oxidase inhibitors
Vinca alkaloids, cisplatin, cyclophosphamide
Disrupt neural pathways
Pulmonary disorders: pneumonia, tuberculosis, fungal infection, empyema, positive pressure ventilation
Central nervous system disorders: infection, trauma, surgery, hemorrhage, inflammatory diseases, degenerative diseases, porphyria
Nephrogenic SIADH
Extremely rare mutation of the V_2 receptor producing chronic activation
Other etiologies
AIDS
Marathon running or other strenuous exercise
Acute psychosis

Specific disorders, which have been reported to cause SIADH, are detailed in Table 5–2, but because of the diffuse regulation of the baroreceptor system and the multiple disorders that can influence vasopressin release via this system, no list should ever be considered complete. SIADH is less common in children than adults and when present is usually chronic.

TREATMENT OF HYPONATREMIA IN SIADH

The therapy of hyponatremia follows pari passu the symptoms of hyponatremia. When hyponatremia is known to have occurred rapidly and is severe and symptomatic, the patient should be treated quickly. When the onset is slow and the syndrome chronic, the treatment should be slow. Acute hyponatremia is arbitrarily defined as having developed over less than 48 hours. Usually, the serum sodium must be less than 120 mEq/L to produce symptomatology. Discontinuing the administration of any hypotonic fluid and the administration of hypertonic NaCl (eg, 3%), possibly with the addition of a diuretic, should be promptly considered. Often acute disorders occur in the hospital in a surgical or obstetrical setting and children and young women are at the greatest risk. For chronic hyponatremia, overly aggressive treatment may cause new and additional pathology. As shown in Figure 5–5 (C), tissue adaptation to chronic hyponatremia results in the loss of intracellular osmolytes.

In the brain these are the same osmolytes that would normally increase in neurons to protect brain volume from a sudden shift of intracellular fluid to the extracellular fluid when there is an acute increase in osmolality of the extracellular fluid. If during the treatment of hyponatremia in SIADH the extracellular fluid osmolality is rapidly increased, the brain has lost the ability to protect intracellular volume and water rapidly moves out of neurons. Brain shrinkage is thought to be the cause of a syndrome of myelinolysis which was first described in the pons—hence the term central pontine myelinolysis. The process can occur diffusely throughout the brain. The cause may be disruption of the blood–brain barrier and influx of plasma components that are toxic to the oligodendrocytes. The syndrome consists of neurologic deterioration over several days with fluctuating consciousness, convulsions, hypoventilation, and hypotension. Eventually, these patients may develop pseudobulbar palsy with difficulty in swallowing and inability to speak, even leading to quadriparesis. Recovery from this syndrome is variable, and many neurologic complications are permanent. MRI scans will demonstrate the demyelination, but usually not until 3 to 4 weeks after correction of the hyponatremia. Chronic alcoholics, patients with malnutrition or liver disease and patients with profound hyponatremia or hypokalemia are at increased risk to develop osmotic demyelination.

In chronic hyponatremia that is asymptomatic, the safest therapy is restriction of free water intake and slow correction of the hyponatremia over days. For chronic hyponatremia that has central nervous system symptoms, carefully controlled and limited increases in the osmolality should be undertaken. Coma or convulsions are obvious signs of neurologic symptoms that may be produced by hyponatremia, but nausea, vomiting, and confusion may be less specific signs of neurologic impairment. There is considerable debate in the literature about the exact rate of correction of the sodium concentration in these cases, and one should review the most up to date recommendations prior to initiating therapy. General parameters of therapy are that the increase of serum sodium might be accomplished at a rate of 0.5 to 1 mEq/L/h with a total correction not to exceed 10–12 mEq/L in the first 24 hours. For those at high risk for osmotic demyelinating the total should not exceed 8 mEq/L in the first 24 hours. In no case should the correction exceed 18 mEq/L in the first 48 hours. Active correction is stopped when the patient's symptoms are abolished, the serum sodium level is greater than 120 mEq/L, or the total magnitude of correction is 18 mEq. Further treatment is undertaken with fluid restriction as for asymptomatic hyponatremia. When overly rapid increases in serum sodium occur, especially in persons at high risk for osmotic demyelination, some experts have given dilute fluids, sometimes with desmopressin, to acutely re-lower the serum sodium. Benefit is supported by some animal studies, but there are no controlled studies documenting that this is beneficial in patients. Fluid intake is restricted in chronic SIADH, but salt intake is not as all of these patients have some degree of sodium loss secondary to the volume expansion and natriuresis.

The tetracycline derivative demeclocycline is known to cause nephrogenic diabetes insipidus and has been used to treat chronic SIADH. Dosages of 600 to 1200 mg/d in divided doses decrease urine osmolality (this use is off-label). Three to four days are required for the effect to reach equilibrium. Azotemia and nephrotoxicity have been reported with demeclocycline especially in patients with cirrhosis. Fluid restriction is the preferred therapy in pediatrics although in young children restricting fluids may not provide adequate calories.

Any of the therapies for hyponatremia and especially fluid restriction will reduce the extracellular fluid volume and, hence, remove the volume-mediated stimulus of renal adaptation illustrated in part D of Figure 5–5 in which the kidney becomes somewhat less responsive to the excess vasopressin by reduction in the amount of aquaporin-2 water channels. When the volume stimulus is removed, there is again an increase in the density of membrane water channels, and the kidney becomes more efficient in retaining water. This accounts for the common clinical observation that fluid restriction that is initially effective may have to be increasingly severe to maintain a beneficial effect.

The most specific treatment for SIADH is to block the V_2 receptors in the kidney. There are recently approved vasopressin receptor antagonists, vaptans, which enhance renal free water excretion (aquaresis) without sodium excretion (natriuresis). Unlike the agonist, desmopressin, these antagonists are nonpeptide drugs. Conivaptan is a combined V_{1A} and V_2 receptor antagonist that is available for intravenous administration to hospitalized patients. Tolvaptan is a selective V_2 receptor antagonist that can be administered orally and is approved to treat severe hyponatremia (serum Na <125 mEq/L) in patients who are symptomatic or patients who have failed to correct with fluid restriction. Vaptans should not be used in conjunction with other therapy, for example, hypertonic saline or fluid restriction. Clinical trials with these agents indicate that they will be useful to treat hyponatremia of SIADH and of other causes. The risk of brain myelolysis caused by too rapid correction of sodium with these agents is the same as with other therapies, so the same recommendations about rates of correction of serum sodium described earlier will apply as physicians gain further experience with these agents. There is little experience with use of vaptans in pediatrics.

SUMMARY

Table 5–3 compares and contrasts pathophysiologic entities associated with decreased vasopressin, diabetes insipidus, and excess vasopressin, SIADH.

OXYTOCIN

The anatomy and synthesis of oxytocin is similar to that described for vasopressin. While there are a number of central nervous system actions that have been attributed to oxytocin (probably acting as a neurotransmitter), the physiologic functions of this posterior pituitary hormone are limited to lactation and parturition.

A. Lactation   The hypothalamic/pituitary hormones critical to lactation are prolactin and oxytocin. Prolactin secretion from the anterior pituitary is described in Chapter 4, and its primary activity is to promote milk production. Oxytocin stimulates milk

TABLE 5–3 Comparison and contrasting the pathophysiology of DI and SIADH.

Pathophysiologic	Lack of Vasopressin: Diabetes Insipidus	Excess Vasopressin: SIADH
Urine	Dilute with increased volume—polyuria	Concentrated with decreased or normal volume
Thirst	Increased and protective—polydipsia	*Not* decreased and *not* protective—normal intake
Serum sodium	Tendency to hypernatremia, but *high normal* due to polydipsia	Decreased—hyponatremia due to water retention
Initial plasma volume	Decreased by polyuria	Increased by water retention
Urine sodium	Variable	Inappropriately elevated for hyponatremia—natriuresis induced by volume expansion
Presentation	Polyuria and polydipsia	Hyponatremia (in SIADH accompanying euvolemia and natriuresis)
Differential diagnostic considerations with the presenting symptoms and laboratory data	*Excess water intake (primary polydipsia where polyuria is normal)* *Solute diuresis producing polyuria and polydipsia, eg, diabetes mellitus*	*Renal Na$^+$ loss (diuretics, Addison disease, renal disease, cerebral salt wasting)* *Elevated urine Na$^+$, but plasma volume is low* *Volume expansion, hyperaldosteronism (CHF, ascites, etc), but urine Na$^+$ is low* *ECF loss (sweating, diarrhea), but volume is decreased and urine Na$^+$ is low*
Decreased or increased synthesis of vasopressin based on anatomy	Tumors with suprasellar origin or extension— infiltrative lesions of the basal hypothalamus—neurohypophyseal infundibulitis	Interruption of inhibitory input from baroreceptors in the chest or CNS
Decreased or increased synthesis based on genetics	Hereditary hypothalamic DI—autosomal recessive AVP gene in AVP producing neurons	Ectopic production of vasopressin by gene derepression in cancer
Disruption of normal control mechanisms	Osmoreceptor system—discrete lesions in anterior hypothalamic area decrease positive signal to AVP neurons and decrease AVP	Baroreceptor system—diffuse lesions in chest, brain stem, and hypothalamus decrease negative input to AVP neurons and increase AVP
Abnormal metabolism of vasopressin	DI in pregnancy	None
Decreased or increased response of V$_2$ receptor	Nephrogenic DI—X-linked recessive mutated inactive V$_2$ receptor	Nephrogenic SIADH—mutated activated V$_2$ receptor—*rare*—unmeasurable AVP
Decreased or increased response of aquaporins	Nephrogenic DI—autosomal recessive mutated inactive aquaporin 2	None
Diagnostic laboratory tests	Elevated pOsm with low uOsm AVP elevated in nephrogenic DI, low in other DI Masked by adrenal or thyroid deficiency, which inhibits free water excretion	Decreased pOsm with high uOsm (not maximally dilute) Urine Na$^+$ inappropriate for hyponatremia AVP elevated in SIADH but also in all differential causes of hyponatremia Normal cortisol, and thyroid and no diuretic therapy; each may cause hyponatremia
Diagnostic imaging—MRI	Head, suprasellar area, tumors, infiltrates, widened stalk, absent posterior pituitary bright spot	Chest—for any pathology, but especially for cancer; head for CNS pathology; full body for cancer
Treatment	Encourage water Diuretics cause volume contraction and reduce urine volume in nephrogenic DI Vasopressin receptor agonists—desmopressin If hypernatremic, overaggressive treatment may rapidly decrease pOsm and produce cerebral edema	Restrict water Drugs that inhibit renal action of vasopressin—demeclocycline Vasopressin receptor antagonists—vaptans Acute onset = rapid therapy Chronic disease = slow therapy Overaggressive treatment to rapidly increase serum Na$^+$ may produce osmotic demyelination

letdown associated with nursing. The milk-producing unit of the breast is the alveolar system in which clusters of milk-producing cells are surrounded by specialized myoepithelial cells. Milk is synthesized in glandular cells of the alveoli. Oxytocin receptors are localized on glandular cells and on myoepithelial cells along the duct. Oxytocin stimulates the cells along the duct to shorten and the ducts to widen, enhancing milk flow through the ducts to the nipple. Suckling at the breast stimulates mechanoreceptors or tactile receptors that ascend through the spinal cord to the lateral cervical nucleus and eventually to the oxytocinergic magnocellular

neurons in the supraoptic and paraventricular nuclei. Neurotransmitters trigger the release of oxytocin. Oxytocin is released in a pulsatile fashion producing a pumping action on the alveoli and promoting maximal emptying of milk from the alveoli. There is no evidence for a central pacemaker to regulate the secretion of oxytocin, and the mechanism of synchrony among individual oxytocinergic neurons is unknown. The importance of oxytocin in maintaining milk secretion is demonstrated in transgenic mice with a knockout of oxytocin synthesis. These animals deliver their young normally, demonstrating the redundant systems for parturition, and produce milk normally, demonstrating the role of prolactin, but are unable to release milk when the pups suckle, demonstrating the importance of oxytocin for milk letdown. The pups die of dehydration with no milk in their stomachs. Administration of oxytocin to the knockout mothers restores milk secretion, and the pups survive.

B. Parturition   Estrogen activates many events that initiate and stimulate parturition whereas progesterone inhibits these events. The uterine myometrial cells have intrinsic contraction activity and are responsive to oxytocin. During pregnancy, oxytocin is released, but oxytocinase decreases the plasma level of oxytocin, and progesterone and relaxin decrease the intrinsic contractility of the myometrium. In humans, there is a dramatic increase in uterine responsiveness to oxytocin as parturition approaches. Several hormones other than oxytocin including prostaglandins, endothelins, adrenergic agonists, corticotropin-releasing hormone, glucocorticoids, and cytokines also participate in the initiation and completion of labor. The most specific role of oxytocin may be the release of oxytocin brought about by vaginal and cervical dilatation produced by the descending head and body, known as the Fergusson reflex. This may be important in stimulating uterine muscle to contract maximally and clamp down blood vessels to decrease blood loss. It is perhaps not surprising that parturition, which is so important to survival of the species, is controlled by many different pathways of cross-stimulation and feed-forward activity. The many redundant systems ensure survival of the species. Oxytocin remains the strongest stimulant to myometrial contraction, explaining its value as a therapeutic agent in inducing parturition and the interest in oxytocin antagonists to delay parturition.

C. Other actions   Other actions of oxytocin have been described in a number of species including effects on behavior of animals, effects on feeding and satiety, response to stress, stimulation of sperm transport, effects on memory, and so forth. None of these have been documented to be physiologically important in humans.

REFERENCES

General References

Morgenthaler NG, Struck J, Jochberger S, Dunser MW. Copeptin: clinical use of a new biomarker. *Trends Endocrinol Metab.* 2008;19:43. [PMID: 18291667]

Muglia LJ, Srivatsa A, Majzoub JA. Disorders of the posterior pituitary. In: Sperling MA, ed. *Pediatric Endocrinology.* 4th ed. 2014:405-443.

Robinson AG, Verbalis JG. In: Melmed S, Polonsky KS, Larsen PR, Kronenberg HM, eds. *Posterior Pituitary in Williams Textbook of Endocrinology.* 13th ed. Philadelphia, PA: Elsevier; 2015:300-332.

Schrier RW, Cadnapaphornchai MA. Renal aquaporin water channels: from molecules to human disease. *Prog Biophys Mol Biol.* 2003;81:117. [PMID: 12565698]

Diabetes Insipidus

Aleksandrov N, Audibert F, Bedard M, Mahone M, Goffinet F, Kadoch I. Gestational diabetes insipidus: a review of an underdiagnosed condition. *J Obstet Gynaecol Can.* 2010;32:225. [PMID: 20500966]

Bichet DG. Vasopressin receptor mutations in nephrogenic diabetes insipidus. *Semin Nephrol.* 2008;28:245. [PMID: 18519085]

Bryant WP, O'Marcaigh AS, Ledger GA, Zimmerman D. Aqueous vasopressin infusion during chemotherapy in patients with diabetes insipidus. *Cancer.* 1994;74:2589. [PMID: 7923015]

Chanson P, Jedynak CP, Dabrowski G, et al. Ultralow doses of vasopressin in the management of diabetes insipidus. *Critical Care Medicine.* 1987;15:44. [PMID: 2431834]

Fujisawa I. Magnetic resonance imaging of the hypothalamic-neurohypophyseal system. *J Neuroendocrinol.* 2004;16:297. [PMID: 15089965]

Kalelioglu I, Kubat Uzum A, Yildirim A, Ozkan T, Gungor F, Has R. Transient gestational diabetes insipidus diagnosed in successive pregnancies: review of pathophysiology, diagnosis, treatment, and management of delivery. *Pituitary.* 2007;10:87. [PMID: 17308961]

Lindheimer MD, Davison JM. Osmoregulation, the secretion of arginine vasopressin and its metabolism during pregnancy. *Eur J Endocrinol.* 1995;132:133. [PMID: 7858729]

Majzoub JA, Srivatsa A. Diabetes insipidus: clinical and basic aspects. *Pediatr Endocrinol Rev.* 2006;4:60. [PMID: 17261971]

Oiso Y, Robertson G, Norgaard JP, Juul KV. Treatment of neurohypophyseal diabetes insipidus. *J Clin Endocrinol Metab.* 2013;98:3958. [PMID: 23884783]

Rupp D, Molitch M. Pituitary stalk lesions. *Curr Opin Endocrinol Diab Obes.* 2008;15:339. [PMID: 18594274]

Sands JM, Bichet DG. Nephrogenic diabetes insipidus. *Ann Intern Med.* 2006;144:186. [PMID: 16461963]

Smith D, Finucane F, Phillips J, et al. Abnormal regulation of thirst and vasopressin secretion following surgery for craniopharyngioma. *Clin Endocrinol (Oxf).* 2004;61:273. [PMID: 15272926]

Smith M. Physiologic changes during brain stem death-lessons for management of the organ donor. *J Heart Lung Transplant.* 2004;9:S217. [PMID: 15381167]

Valenti G, Procino G, Tamma G, Carmosino M, Svelto M. Minireview: aquaporin 2 trafficking. *Endocrinology.* 2005;146:5063. [PMID: 16150901]

Van de Walle J, Stockner M, Raes A, Norgaard JP. Desmopressin 30 years in clinical use: a safety review. *Curr Drug Safety.* 2007;2(3):232. [PMID: 18690973]

SIADH

Bartter FC, Schwartz WB. The syndrome of inappropriate secretion of antidiuretic hormone. *Am J Med.* 1967;42:790. [PMID: 5337379]

Berl T. Vasopressin antagonists. *N Engl J Med.* 2015;372:2207. [PMID: 26039601]

Ecelbarger CA, Murase T, Tian Y, Nielsen S, Knepper MA, Verbalis JG. Regulation of renal salt and water transporters during vasopressin escape. *Prog Brain Res.* 2002;139:75. [PMID: 12436927]

Feldman BJ, Rosenthal SM, Vargas GA, et al. Nephrogenic syndrome of inappropriate antidiuresis. *N Engl J Med.* 2005;352:1884. [PMID: 15872203]

Greenberg A, Verbalis JG. Vasopressin receptor antagonists. *Kidney Int.* 2006;69:2124. [PMID: 16672911]

Kratz A, Siegel AJ, Verbalis JG, et al. Sodium status of collapsed marathon runners. *Arch Pathol Lab Med.* 2005;129:227. [PMID: 15679427]

Loh JA, Verbalis JG. Disorders of water and salt metabolism associated with pituitary disease. *Endocrinol Metab Clin North Am.* 2008;37:213. [PMID: 18226738]

Sterns RH, Baer J, Ebersol S, Thomas D, Lohr JW, Kamm DE. Organic osmolytes in acute hyponatremia. *Am J Physiol.* 1993;264:F833. [PMID: 8498536]

Sterns RH, Thomas DJ, Herndon RM. Brain dehydration and neurologic deterioration after rapid correction of hyponatremia. *Kidney Int.* 1989;35:69. [PMID: 2709662]

Verbalis JG, Goldsmith SR, Greenberg A, et al. Diagnosis, evaluation, and treatment of hyponatremia: expert panel recommendations. *Am J Med.* 2013; 126:S1. [PMID: 24074529]

Verbalis JG, Murase T, Ecelbarger CA, Nielsen S, Knepper MA. Studies of renal aquaporin-2 expression during renal escape from vasopressin-induced antidiuresis. *Adv Exp Med Biol.* 1998;449:395. [PMID: 10026831]

Cerebral Salt Wasting

Berkenbosch JW, Lentz CW, Jimenez, DF, Tobias JD. Cerebral salt wasting syndrome following brain injury in three pediatric patients: suggestions for rapid diagnosis and therapy. *Pediatr Neurosurg.* 2002;36:75. [PMID: 11893888]

Betjes MG. Hyponatremia in acute brain disease: the cerebral salt wasting syndrome. *Eur J Intern Med.* 2002;13:9. [PMID: 11836078]

Cerda-Esteve M, Cuadrado-Godia E, Chillaron JJ, et al. Cerebral salt wasting syndrome: review. *Eur J Intern Med.* 2008;19:249. [PMID: 18471672]

Donati-Genet PC, Dubuis JM, Girardin E, Rimensberger PC. Acute symptomatic hyponatremia and cerebral salt wasting after head injury: an important clinical entity. *J Pediatr Surg.* 2001;36:1094. [PMID: 11431791]

Palmer BF. Hyponatremia in patients with central nervous system disease: SIADH versus CSW. *Trends Endocrinol Metab.* 2003;14:182. [PMID: 12714279]

Sviri GE, Feinsod M, Soustiel JF. Brain natriuretic peptide and cerebral vasospasm in subarachnoid hemorrhage. Clinical and TCD correlations. *Stroke.* 2000;31:118. [PMID: 10625725]

von Bismarck P, Ankermann T, Eggert P, Claviez A, Fritsch MJ, Krause MF. Diagnosis and management of cerebral salt wasting (CSW) in children: the role of atrial natriuretic peptide (ANP) and brain natriuretic peptide (BNP). *Childs Nerv Syst.* 2006;22:1275. [PMID: 16607534]

Lactation

Buhimschi CS. Endocrinology of lactation. *Obstet Gynecol Clin North Am.* 2004;31:963. [PMID: 15550345]

Neville MC, Morton J, Umemura S. The transition from pregnancy to lactation. *Pediatr Clin North Am.* 2001;48:35. [PMID: 11236732]

Parturition

Blanks AM, Thornton S. The role of oxytocin in parturition. *Br J Obstet Gyn.* 2003;110(Suppl 20):46. [PMID: 12763111]

Brown AG, Leite RS, Strauss JF, 3rd. Mechanisms underlying "functional" progesterone withdrawal at parturition. *Ann NY Acad Sci.* 2004;1034:36. [PMID: 15731298]

Gross G, Imamura T, Muglia LJ. Gene knockout mice in the study of parturition. *J Soc Gynecol Investig.* 2000;7:88. [PMID: 10785607]

Hertelendy F, Zakar T. Prostaglandins and the myometrium and cervix. *Prostaglandins Leukot Essent Fatty Acids.* 2004;70:207. [PMID: 14683694]

Smith JG, Merrill DC. Oxytocin for induction of labor. *Clin Obstet Gynecol.* 2006;49:594. [PMID: 16885666]

Smith R. Parturition. *N Engl J Med.* 2007;356:271. [PMID: 17229954]

Young WS, 3rd, Shepard E, Amico J, et al. Deficiency in mouse oxytocin prevents milk ejection, but not fertility or parturition. *J Neuroendocrinol.* 1996;8:847. [PMID: 8933362]

Other Oxytocin

Marazziti D, Catena Dell'osso M. The role of oxytocin in neuropsychiatric disorders. *Curr Med Chem.* 2008;15:698. [PMID: 18336283]

This page intentionally left blank

Growth

Dennis Styne, MD

ACTH	Adrenocorticotropin hormone		**IUGR**	Intrauterine growth retardation or restriction
ALS	Acid labile subunit		**JAK-STAT**	Janus kinase-signal transducers and activators of transcription
cAMP	Cyclic adenosine monophosphate			
CPHD	Combined pituitary hormone deficiency		**LDL**	Low-density lipoprotein
EGF	Epidermal growth factor		**LH**	Luteinizing hormone
EGF-R	Epidermal growth factor receptor		**LS**	Lower segment
FGF	Fibroblast growth factor		**MC4R**	Melanocortin 4 receptor
FGF-R	Fibroblast growth factor receptor		**NSD1**	Nuclear receptor–binding SET domain protein 1
GH	Growth hormone			
GHBP	Growth hormone–binding protein		**Pit-1**	Pituitary transcription factor 1
GHRH	Growth hormone–releasing hormone		**Prop-1**	Prophet of Pit-1 homeobox 1
GHSR	Growth hormone secretagogue receptor		**PTH**	Parathyroid hormone
GnRH	Gonadotropin-releasing hormone		**PTPN11**	Protein tyrosine phosphatase nonreceptor 11
hCG	Human chorionic gonadotropin			
hCS	Human chorionic somatomammotropin		**RTA**	Renal tubular acidosis
HDL	High-density lipoprotein		**SGA**	Small-for-gestational age
HESX1	Hesx1 homeodomain		**SHOX**	Short stature homeobox
hGH	Human growth hormone		**SOD**	Septo-optic dysplasia
IGF-I	Insulin-like growth factor I		**SRIF**	Somatostatin
IGF-II	Insulin-like growth factor II		**TBG**	Thyroxine-binding globulin
IGFBP	Insulin-like growth factor–binding protein		**TRH**	Thyrotropin-releasing hormone
			TSH	Thyrotropin
IGHD	Isolated growth hormone deficiency		**US**	Upper segment

Assessment of growth in stature is an essential part of the pediatric examination. Growth is an important index of physical and mental health and of the quality of the child's psychosocial environment; chronic problems in any of these areas may be reflected in a decreased growth rate which may be a critical clue as to the age of onset of the condition. We shall consider influences on normal growth, the normal growth pattern, the measurement of growth, and conditions that lead to disorders of growth.

NORMAL GROWTH

INTRAUTERINE GROWTH

The growth of a fetus begins with a single fertilized cell and ends with differentiation into more than 200 cell types, length increasing by 5000-fold, surface area by 6×10^6-fold, and weight by 6×10^{12}-fold. This all to leads to an approximately 7 lb newborn.

Overall, the growth of the fetus is dependent on the availability of adequate oxygen and nutrition delivered by the placenta and is orchestrated by a group of growth factors, all overseen by a basic genetic plan. Genetic factors are more important early in gestation, whereas the maternal environment attains more importance late in gestation.

The classic definition of small-for-gestational age (SGA), is a birth weight two standard deviations (SDs) below the mean or below the 5th percentile for birth weight, or birth weight below 2500 g for a term infant in the United States. The term intrauterine growth retardation or intrauterine growth restriction (IUGR) is not synonymous with SGA, because IUGR refers to decreased intrauterine growth velocity noted on ultrasound. Statistics and charts showing various percentiles of weight for gestational age are available to determine which premature infants are SGA and what weights are appropriate for gestational age (AGA). About 20% of SGA infants remain short as children and adults, in contrast to appropriate-for-gestational-age premature infants, who are smaller at birth but generally experience catch-up growth in the first 2 years. Those SGA infants who do not experience catch-up growth by 2 years of age may be candidates for growth hormone (GH) therapy. Recent studies suggest that preterm, AGA babies who show poor growth in the first three postnatal months follow growth patterns similar to SGA infants; thus consideration of GH therapy may be warranted as this situation receives further study.

THE PLACENTA

The placenta influences most aspects of fetal growth, including the supply of adequate nutrition and oxygen and also acts as an endocrine organ by producing and releasing hormones and growth factors. Aberrant delivery or control of any of these factors affects fetal growth. Placental weight is usually directly related to birth weight.

CLASSIC HORMONES OF GROWTH AND FETAL GROWTH

The hormones that mediate postnatal growth do not necessarily play the same roles in fetal growth. GH is present in high concentrations in the fetus, in contrast to the limited presence of GH receptors. Although this discrepancy suggests limited activity of GH in the fetus, GH does play a role in fetal growth as reflected in the average birth weight at 1 SD below the mean in GH-deficient infants. Infants with GH resistance due to abnormal, reduced or absent GH receptors (eg, Laron syndrome) have elevated GH and low serum insulin-like growth factor (IGF)-I levels; they also have decreased birth length and weight. Thyroid hormone deficiency does not directly affect human birth weight, but prolonged gestation can be a feature of congenital hypothyroidism, and this factor will itself increase weight. Placental lactogen exerts no effect on birth size in human beings. However, the concentration of placental-derived GH (from the *GHV* gene) is significantly decreased in the serum of a pregnant woman bearing a fetus with IUGR.

GROWTH FACTORS AND ONCOGENES IN FETAL GROWTH

Oncogenes may be responsible for neoplastic growth in postnatal life, but expression of these genes is important in the normal development of many fetal organs. Remarkably, the same oncogenes that cause postnatal neoplasia are prevented from causing tumors in the normally differentiating fetus. For example, a mutation in the *von Hippel-Lindau* gene, normally acting as tumor suppressor gene, predisposes to retinal, cerebellar, and spinal hemangioblastomas, renal cell carcinomas, and pheochromocytomas, but the normal *VHL* gene is expressed in all three germ cell layers of the embryo and in the central nervous system (CNS), kidneys, testis, and lung of the fetus, playing a role in normal fetal development for this gene.

INSULIN-LIKE GROWTH FACTORS, RECEPTORS, AND BINDING PROTEINS

IGF-I in the fetus is regulated by metabolic factors other than GH, in contrast to the dependence of IGF-I generation upon GH in postnatal life. One explanation is that there are fewer GH receptors in the fetus than after birth. In the human fetus, serum GH falls during later gestation owing to maturation of central nervous system's negative control, whereas serum IGF-I and IGF-binding protein (IGFBP)-3 rise during gestation, demonstrating their independence from GH stimulation.

Studies of knockout mice, which lack various growth factors or binding proteins, indicate a role for IGF-II in growth during early gestation and one for IGF-I during later gestation. Knockout of type 1 IGF receptors leads to a more profound growth failure than is found in IGF-I knockout mice alone, suggesting that factors other than IGF-I (eg, IGF-II) exert effects on fetal growth through the type 1 receptor.

Study of transgenic mice overexpressing IGFBPs supports the concepts that excess IGFBP-1 stunts fetal growth while excess IGFBP-3 leads to selective organomegaly. For example, overexpression of IGFBP-3 in mice led to organomegaly of the spleen, liver, and heart, although birth weight was not different from that of wild-type mice.

Although controversy remains over some of the data regarding IGFs and fetal growth, a summary of the complex IGF system in the fetus, based on the evidence from various species, appears to apply to the human being as follows.

1. IGFs are detectable in many fetal tissues from the first trimester onward.
2. Concentrations of IGFs in the fetal circulation increase during pregnancy, and at term the concentrations of IGF-I are directly related to birth weight.
3. In mice, disruption of the *IGF* genes leads to severe growth retardation.
4. There is a striking increase in IGFBP-1 and IGFBP-2 concentrations in amniotic fluid at the end of the first trimester.
5. The major binding proteins in the human fetus are IGFBP-1 and IGFBP-2.

6. From as early as 16 weeks, there is an inverse correlation between fetal concentrations of IGFBP-1 and birth weight.
7. In the mother, circulating concentrations of IGF-I and IGFBP-1 increase during pregnancy.
8. Maternal IGFBP-1 concentrations are elevated in severe pre-eclampsia and IUGR.
9. Fetal concentrations of IGFBP-1 are elevated in cases of IUGR, especially those associated with specific evidence of reduced uteroplacental blood flow. Production of IGFBP-1 appears to be a sensitive indicator of the short- or long-term response to reduced fetal nutrition.

INSULIN

Although insulin is a major regulatory factor for carbohydrate metabolism, many lines of evidence demonstrate that it is a growth factor as well and has importance in fetal growth. Macrosomia is a well-known effect of fetal hyperinsulinism as is found in the infant of the diabetic mother. Increased weight gain in pregnant women over 40 lb leads to significantly increased risk of fetal macrosomia in gestational diabetes mellitus as well as in those with normal glucose tolerance test results. Errors in the normal pattern of *IGF-II* gene expression from the paternal chromosome and type 2 IGF receptor (for IGF-II) from the maternally derived gene underlie the pathogenesis of Beckwith-Wiedemann syndrome (OMIM #130650) a autosomal dominant disorder with mutation, deletion or hypermethylation of genes at 11p15.5. Affected infants are large and have elevated insulin concentrations; characteristics include exomphalos, macroglossia, and gigantism while hepatoblastoma and Wilms tumor may occur as as well.

Just as increased insulin stimulates fetal growth, syndromes of fetal insulin deficiency, such as congenital diabetes mellitus, pancreatic dysgenesis, or fetal insulin resistance (eg, leprechaunism), are characterized by IUGR. Infants born to diabetic mothers with vascular disease, hypertension, and/or eclampsia or pre-eclampsia also have IUGR and are born SGA. In that situation, it is clear that limited nutrient delivery compromises the growth of the infant.

EPIDERMAL GROWTH FACTOR

Epidermal growth factor (EGF) is involved with fetal growth, and expression varies with disordered fetal growth. Microvilli purified from the placentas of infants with IUGR have decreased or absent placental EGF receptor (EGF-R) phosphorylation and tyrosine kinase activity. Maternal smoking decreases birth weight by an average of 200 g, with the major effect occurring late in pregnancy; the placenta responds to smoking by significant changes in its vascularity, which leads to fetal hypoxia. There are decreased numbers of EGF-Rs and a reduced affinity of these receptors for EGF in the placentas of women who are smokers. Hypertensive patients also have decreased numbers of placental EGF-Rs, which may result in IUGR.

EGF levels in amniotic fluid are normally increased near term but decreased in pregnancies complicated by IUGR—although not, conversely, increased in infants who are large for gestational age.

EGF levels in the first urines to be voided by IUGR and macrosomic infants are lower than in control infants.

EGF administered to fetal monkeys results in histologic and biochemical maturation of their lungs, leading to improved air exchange and a diminished requirement for respiratory support. Surfactant apoprotein-A concentration and the lecithin-sphingomyelin ratio are both significantly higher in the amniotic fluid of EGF-treated fetuses. Whereas birth weight is not affected by EGF, adrenal and gastrointestinal (GI) system weights, standardized for body weight, are increased significantly. Furthermore, EGF stimulates GI muscle, GI enzyme maturation, and GI size improving the ability of the infant monkey to absorb nutrients. Lastly, EGF advances the maturation of the fetal adrenal cortex, increasing the expression of 3-beta-hydroxysteroid dehydrogenase.

FIBROBLAST GROWTH FACTOR

Genetically engineered fibroblast growth factor receptor 2 (FGF-R)-deficient mice are severely growth-retarded and die before gastrulation. Aberrant FGF signaling during limb and skeletal development in humans can lead to syndromes of dysmorphia. For example, achondroplasia (OMIM #100800) is due to mutations in the transmembrane domain of the type 3 FGF-R.

GENETIC, MATERNAL, AND UTERINE FACTORS

Maternal factors, often expressed through the uterine environment, exert more influence on birth size than paternal factors. The height of the mother correlates better with fetal size than the height of the father. However, there is a genetic component to length at birth that is not sex specific. Firstborn infants are on the average 100 g heavier than subsequent infants; maternal age over 38 years leads to decreased birth weight; and male infants are heavier than female infants by an average of 150 to 200 g. Poor maternal nutrition is the most important condition leading to low birth weight and length on a worldwide basis. Chronic maternal disease and eclampsia can also lead to poor fetal growth. Maternal alcohol ingestion has severe adverse effects on fetal length and mental development and predisposes to other physical abnormalities seen in the fetal alcohol syndrome; these include microcephaly, mental retardation, midfacial hypoplasia, short palpebral fissures, wide-bridged nose, long philtrum, and narrow vermilion border of the lips. Affected infants never recover from this loss of length but attain normal growth rates in the postnatal period. Abuse of other substances and chronic use of some medications (eg, phenytoin) can cause IUGR. Cigarette smoking causes not only retarded intrauterine growth but also decreased postnatal growth for as long as 5 years after parturition. Maternal infection—most commonly toxoplasmosis, rubella, cytomegalovirus infection, herpes simplex infection, and HIV infection—leads to many developmental abnormalities as well as short birth length. In multiple births, the weight of each fetus is usually less than that of the average singleton. Uterine tumors or malformations may decrease fetal growth.

CHROMOSOMAL ABNORMALITIES AND MALFORMATION SYNDROMES

Infants with abnormal karyotypes may have malformation syndromes and may also demonstrate poor fetal or postnatal growth. In most cases, endocrine abnormalities have not been noted. For further discussion of this extensive subject, the reader is referred to other sources listed in the references at the end of this chapter.

FETAL ORIGINS OF ADULT DISEASE

The metabolic syndrome consists of (1) hypertension, (2) impaired glucose tolerance, and (3) elevated triglycerides among other features (see Chapter 17). Insulin resistance is a cardinal feature and might be the basis for most or all of these complications or may be just one feature of the syndrome, according to some. The metabolic syndrome is one of the long-lasting effects of abnormalities in fetal growth. Evidence from many international studies indicates a relationship between low birth weight or low weight at 1 year of age and chronic disease in adulthood. An opposing view is that catch-up growth, rather than low birth weight, is responsible for these defects long into the child's future.

Inanition during the last two trimesters of pregnancy, which occurred during the famine in Holland during World War II, led to an 8% to 9% decrease in birth weight. On the other hand, in the Dutch famine, infants born after early gestational starvation of their mothers, but with improved maternal nutrition in late gestation were of normal size at birth. However, the female infants, born of normal size after this early gestational maternal starvation, themselves gave birth to small infants (SGA of 300-500 g decrease). In other populations, women with a history of SGA tend to have SGA infants themselves, and some studies show that generations of malnutrition must be followed by generations of normal nutrition before there is correction of the birth weight of subsequent infants. The sparse environment of the fetus early in gestation in a mother with various degrees of starvation *programs* the fetal metabolism for survival of the fetus but later in life these survival techniques become maladaptive in an environment of plenty. Insulin resistance in fetal life may spare nutrients from utilization in muscle, thus leaving them available for the brain; this mechanism would serve to minimize CNS damage in the fetus during periods of malnutrition. The complexity of the situation is not completely understood but is the target of extensive research *in vivo*, *in vitro*, and in long-term clinical studies.

Birth weight and rate of postnatal growth (ie, catch-up growth)—not prematurity alone—are inversely related to cardiovascular mortality and the prevalence of the metabolic syndrome. There is concern when a premature or SGA infant fails to gain weight after birth approximating the normal growth curve. Efforts are made to increase the growth rate often to the supramaximal velocity. Unfortunately, excessive growth soon after birth can also lead to obesity and its comorbidities later in life. It is not yet clear what the optimal nutrition is for a premature or SGA infant to avoid this metabolic programming. Further research as needed to determine the optimal growth rate for newborns experiencing poor intrauterine growth or premature delivery.

Studies of otherwise normal but thin children, who had a history of SGA, demonstrated insulin resistance before the teenage years, supporting the concept of early metabolic programming. Present-day adults, who were born in the Netherlands during the Dutch famine who had the lowest birth weights and the lowest maternal weights (those subjects whose mothers experienced malnutrition during the last two trimesters noted earlier), show a degree of insulin resistance that is directly related to their degree of SGA, further documenting the relationship between fetal undernutrition and adult insulin resistance as the etiology of poor growth.

On the other hand, macrosomic infants born to mothers with diabetes mellitus often develop childhood obesity and insulin resistance in later childhood even if they have a period of normal weight between 1 and 5 years of age. Remarkably, studies of the offspring of Dutch mothers exposed to famine during World War II in the first two trimesters (the time in which birth weight is least affected by maternal starvation) demonstrated a twofold increase in the incidence of obesity at 18 years of age, compared with a 40% decrease in the incidence of obesity if the individual was exposed to famine in the last trimester (the time in which birth weight is most negatively impacted by maternal starvation).

Individuals born SGA also have variations in pubertal development and reproductive hormones. SGA girls tend to have earlier puberty, or if puberty occurs at an average age, more rapid progression of puberty or polycystic ovarian syndrome (PCOS). Adult males born SGA have increased aromatase and increased 5-alpha reductase demonstrating effects on reproduction in the male as well as the female. In sum, prenatal and early postnatal growth affects the older child and adult in many ways.

POSTNATAL GROWTH

Postnatal growth in stature follows a characteristic pattern in normal children (Figures 6–1 and 6–2). The highest overall growth rate occurs in the fetus, the highest postnatal growth rate just after birth, and a slower growth rate follows in mid-childhood (Figures 6–3 and 6–4). There are two periods characterized by brief growth spurts in childhood: the infant-childhood growth spurt between 1½ years and 3 years and the mid-childhood growth spurt between 4 and 8 years. The mid-childhood growth spurt does not occur in all children, is more frequent in boys than girls, and its presence is hereditary. After another plateau, the striking increase in stature, the pubertal growth spurt follows, causing a second peak growth velocity. The final decrease in growth rate then ensues, until the epiphyses of the long bones fuse and growth ceases.

In addition, an *adiposity rebound* of accelerating weight gain and rising body mass index (BMI) occurs in early childhood after a period of relative stability of weight gain. An early adiposity rebound is a risk factor for the development of obesity later in childhood and thereafter.
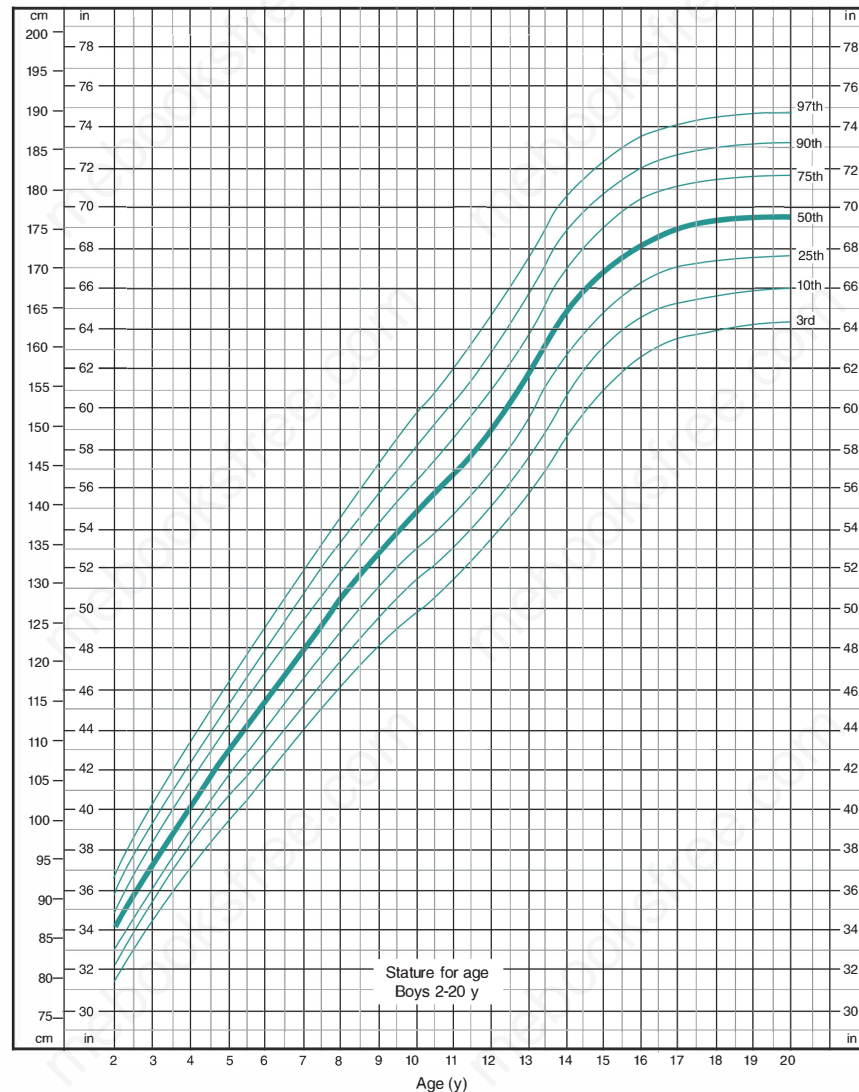
FIGURE 6–1 Growth chart for boys in most clinical situations showing the 3rd to the 97th percentiles. The lines for the 3rd and the 97th percentiles approximate those for −2.5 SDs and +2.5 SDs from the mean. (Reproduced with permission from National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion. http://www.cdc.gov/growthcharts. Published May 30, 2000.)

Endocrine Factors

A. Growth hormone and insulin-like growth factors   As discussed in Chapter 4, somatotropin or GH is suppressed by hypothalamic GH release-inhibiting factor (somatostatin or SRIF) and stimulated by GH-releasing hormone (GHRH or GRF). The gene for GH is located on the long arm of chromosome 17 in a cluster of five genes: *GHN* codes for human GH (a single 191-amino-acid polypeptide chain with a molecular weight of 22 kD); *GHV* codes for a variant GH produced in the placenta; *CSH1* and *CSH2* codes for prolactin; and *CSHP1* codes for a variant prolactin molecule. A 20-kD variant of pituitary GH accounts for 5% to 10% of circulating GH and is derived from the same gene as the 22K variant, *GHN*, but results from alternative splicing. The 20-kD variant is poorly characterized but when derived from the

placental gene, *GHV*, *in vitro* it has less diabetogenic effect but similar growth-promoting and lipolytic activity.

The effects of GH are mainly mediated by the IGFs, but GH also directly stimulates lipolysis, increased amino acid transport into tissues, and increased protein and glucose synthesis in liver. It also has direct effect on cartilage growth. GH in excess causes insulin resistance and is diabetogenic. GH is secreted in a pulsatile manner, so that serum concentrations are low much of the day, but peak during short intervals. Values are higher in the immediate neonatal period, decrease through childhood, and rise again as a result of increased pulse amplitude (but not frequency) during puberty. GH secretion falls again during aging.

GH circulates in plasma bound to a protein, the GH-binding protein (GHBP), with a sequence equivalent to that of the
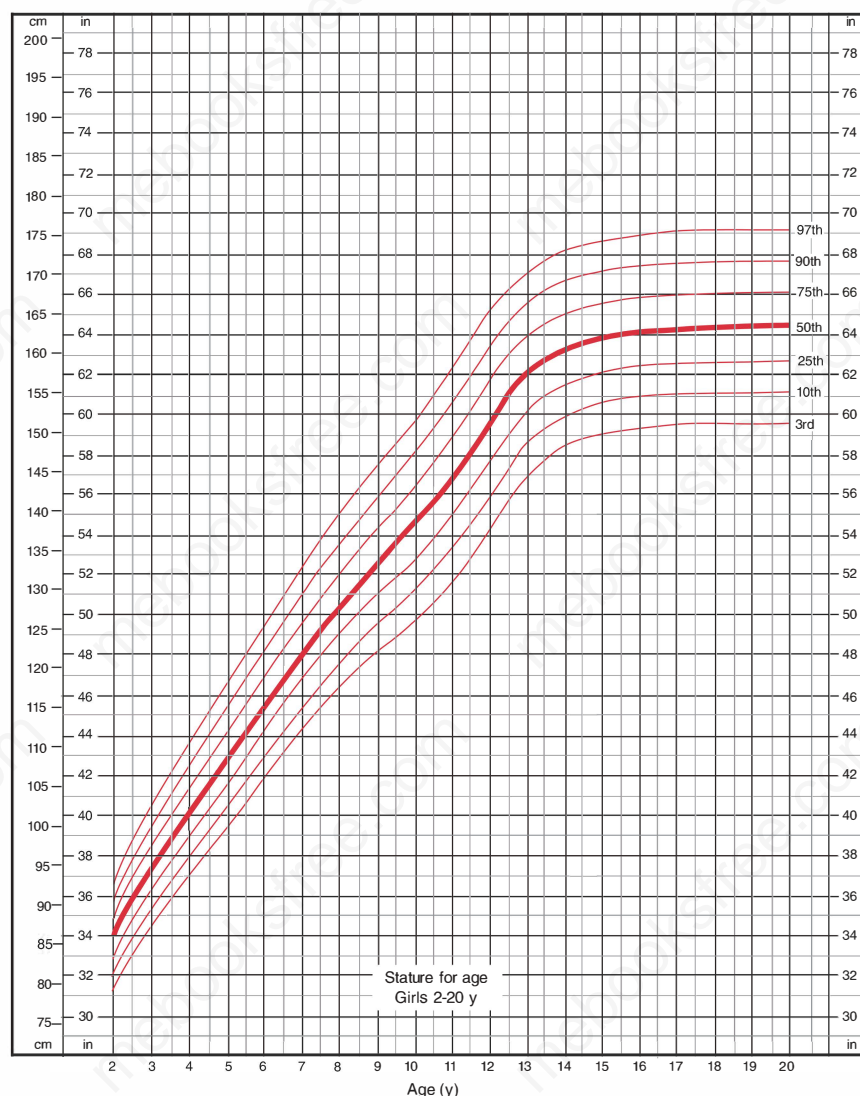
FIGURE 6–2 Growth chart for girls in most clinical situations showing the 3rd to 97th percentiles. The lines for the 3rd and the 97th percentiles approximate those for −2.5 SDs and +2.5 SDs from the mean. (Reproduced with permission from National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion. http://www.cdc.gov/growthcharts. Published May 30, 2000.)

extracellular membrane domain of the GH receptor from which it is cleaved. The physiology of the GHBP appears to reflect important interrelationships between GH and GHR in terms of effects on growth. For example, obese patients have lower plasma GH concentrations but higher GHBP levels, whereas starvation raises GH concentrations and lowers GHBP levels. Patients with abnormalities of the GH receptor (eg, Laron dwarfism) also have a defect reflected in the serum GHBP concentrations; those with decreased numbers of GH receptors have decreased serum GHBP concentrations. Patients who are unable to dimerize the GH receptors to allow activation of the GH-GHR complex or those who have intracellular defects in the JAK-STAT system have no alteration in GHBP concentration, but are short (see Chapters 1 and 4).

GH exerts its effects on growth mainly, but not solely, through the IGFs and their binding proteins. IGF-I and IGF-II have

structures similar to that of the proinsulin molecule but differ from insulin in regulation, receptors, and biologic effects. The IGFs were originally called sulfation factor and then somatomedin. The single copy gene for pre-pro-IGF-I is located on the long arm of chromosome 12. Post-translational processing produces the 70-amino acid mature form; alternative splicing mechanisms produce structural variants of the molecule. The IGF-I cell membrane receptor (the type 1 receptor) resembles the insulin receptor in its structure consisting of two alpha and two beta chains. Binding of IGF-I to type 1 receptors stimulates tyrosine kinase activity and autophosphorylation of tyrosine residues in the receptor. This leads to cell differentiation or division (or both). IGF-I receptors are downregulated by increased IGF-I concentrations, whereas decreased IGF-I concentrations increase IGF-I receptors.

IGF molecules in the circulation are mostly bound to one of eight IGFBPs. IGFBP-1 and IGFBP-3 have been most extensively
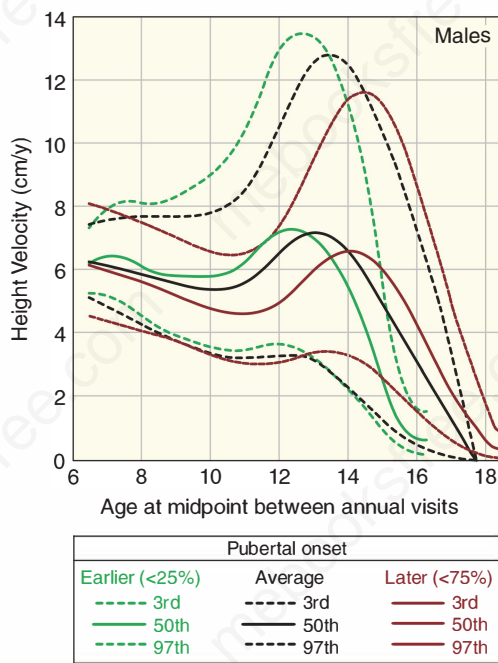
FIGURE 6–3 Reference curves for height velocity for boys aged 6 to 19 years according to onset of puberty. The median and the 3rd and 97th percentiles for earlier onset puberty is in blue, for average onset is in black, and for later onset is in red. Earlier puberty was defined as age of onset 10.2 years, average onset as occurring between 10.2 and 11.8 years, and later puberty as onset at age >11.8 years. The charts can be used clinically by calculating the yearly increment in stature as the difference between measurements of height at 2 times $(g_2 - g_1)$ divided by the difference of the age of the child at the first measurement subtracted from the age at the second measurement, $(t_2 - t_1)$. The quotient of this division is plotted on the chart as a bar which starts at the initial age of the child and ends at the age of the child at the end of the interval with the center of the bar plotted at the age $(t_1 + t_2)/2$. (Reproduced with permission from Kelly A, Winer KK, Kalkwarf H. Age-based reference ranges for annual height velocity in US children. *J Clin Endocrinol Metab*. 2014 Jun;99(6):2104-2112.)

FIGURE 6–4 Reference curves for height velocity for girls aged 6 to 17 years according to onset of puberty. The median and the 3rd and 97th percentiles for earlier onset puberty is in blue, for average onset is in black, and for later onset is in red. Earlier puberty was defined as age of onset 10.2 years, average onset as occurring between 10.2 and 11.8 years, and later puberty as onset at age >11.8 years. The charts can be used clinically by calculating the yearly increment in stature as the difference between measurements of height at 2 times $(g_2 - g_1)$ divided by the difference of the age of the child at the first measurement subtracted from the age at the second measurement $(t_2 - t_1)$. The quotient of this division is plotted on the chart as a bar which starts at the initial age of the child and ends at the age of the child at the end of the interval with the center of the bar plotted at the age $(t_1 + t_2)/2$. (Reproduced with permission from Kelly A, Winer KK, Kalkwarf H. Age-based reference ranges for annual height velocity in US children. *J Clin Endocrinol Metab*. 2014 Jun;99(6):2104-2112.)

studied. IGFBP-1 is a 25-kD protein and mainly inhibits IGF-I action. Serum levels of IGFBP-1 are inversely proportional to insulin levels. This protein does not appear to be regulated by GH. It is present in high concentrations in fetal serum and amniotic fluid. Newborn serum concentrations of IGFBP-1 are inversely proportional to birth weight.

IGF-I circulates bound to IGFBP-3 and an acid labile subunit (ALS) in a 150-kD complex. Abnormalities of the ALS lead to decreased growth. Serum IGFBP-3 concentrations are directly proportional to GH concentrations but also to nutritional status. In malnutrition, IGFBP-3 and IGF-I levels fall, while GH rises as does IGFBP-1. IGF-I directly regulates IGFBP-3 as well. IGFBP-3 rises with advancing age through childhood, with highest values achieved during puberty; however, the pattern of change of IGF-I at puberty is different from that of IGFBP-3. The molar ratio of IGF-I to IGFBP-3 rises at puberty, suggesting that more IGF-I is free to influence growth during this period.

IGF-I is produced in most tissues and is exported to neighboring cells to act on them in a paracrine manner or on the cell of origin in an autocrine manner. Thus, serum IGF-I concentrations may not reflect the most significant actions of this growth factor. The liver is a major site of IGF-I synthesis, and much of the circulating IGF-I probably originates in the liver; serum IGF-I concentrations vary in liver disease, falling with reductions in functional hepatic mass. IGF-I is a progression factor, so that a cell which has been exposed to a competence factor such as platelet-derived growth factor (PDGF) in stage G_0 of the cell cycle and has progressed to G_1 can, with IGF-I exposure in G_1, undergo division in the S phase of the cell cycle. Aside from the stimulatory effects of IGF-I on cartilage growth, IGF-I has stimulatory effects on hematopoiesis, ovarian steroidogenesis, myoblast proliferation and differentiation, and differentiation of the lens.

IGF-I is now available in recombinant form. IGF-I administration increases nitrogen retention and decreases blood urea nitrogen. While GH is diabetogenic, IGF-I can lower blood sugar values. In GH-resistant patients (Laron dwarfs), IGF-I stimulates growth in the absence of GH. IGF-I is approved by the Food and Drug Administration (FDA) for use in short stature due to primary IGF-I deficiency. IGF-II is a 67-amino acid peptide. The gene for pre-pro-IGF-II is located on the short arm of chromosome 11, close to the gene for preproinsulin. The type 2 IGF receptor (cation-independent mannose 6-phosphate receptor) preferentially binds IGF-II and is identical to the mannose 6-phosphate receptor, a single-chain transmembrane protein. Although most of the effects of IGF-II appear to be mediated by its interaction with the type 1 receptor, independent actions of IGF-II, via the type 2 receptor, have been described.

Plasma concentrations of the IGFs vary with age and physiologic condition. IGF-I concentrations are low at term in neonates and remain relatively low in childhood until a peak is reached during puberty, with values rising higher than at any other time in life. Serum IGF-I then decreases to adult levels, values higher than in childhood but lower than in puberty. With advancing age, serum GH and IGF-I levels decrease. IGF-I concentrations are more highly correlated in monozygotic twins than in same-sex dizygotic twins, indicating a genetic effect on regulation of IGF-I levels.

GH deficiency leads to lower serum IGF-I and IGF-II concentrations, whereas GH excess leads to elevated IGF-I but no rise in IGF-II above normal. Because serum IGF-I is lower during states of nutritional deficiency, IGF-I is not a perfect tool in the differential diagnosis of conditions of poor growth, which often include impaired nutritional state. IGF-I suppresses GH secretion by a negative feedback mechanism, so that patients who lack GH receptors (Laron dwarfs) or who are unable to produce IGF-I, have elevated GH concentrations but negligible IGF-I concentrations. Rare patients with poor growth who lack IGF-I receptors have elevated concentrations of IGF-I which exerts no biological activity.

B. Thyroid hormone   As noted earlier, newborns with congenital hypothyroidism are of normal length, but if untreated, they manifest exceedingly poor growth soon after birth. Infants with untreated congenital hypothyroidism suffer permanent developmental delay so early treatment is necessary. Newborn screening for congenital hypothyroidism is universal in the United States and most countries. Acquired hypothyroidism with onset after 3 years leads to a markedly decreased growth rate but no permanent intellectual defects. Bone age advancement is severely delayed in hypothyroidism, usually more so than in GH deficiency, and epiphyseal dysgenesis is seen as calcification of the epiphyses progresses. The normal decrease in the upper to lower segment ratio (U/L ratio) with age (Figure 6–5) is delayed, and therefore, the U/L ratio is elevated, owing to poor lower limb growth in hypothyroidism.

C. Sex steroids   Gonadal sex steroids exert an important influence on the pubertal growth spurt, whereas absence of these
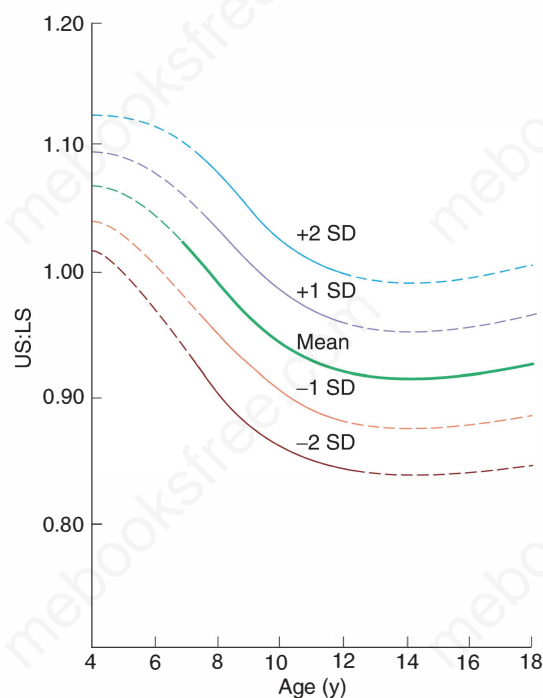
FIGURE 6–5 Normal upper:lower segment (US:LS) ratios, based on findings in 1015 white children. Values are slightly lower for black children. (Reproduced with permission from McKusick V. *Hereditable Disorders of Connective Tissue.* 4th ed. St. Louis: Mosby; 1972.)

factors is not of major importance in prepubertal growth. Gonadal and adrenal sex steroids in excess can cause a sharp increase in growth rate as well as the premature appearance and progression of secondary sexual features. If unabated, increased sex steroids will cause advancement of skeletal age, premature epiphyseal fusion, and short adult stature all mediated by estrogen. The pubertal rise in gonadal steroids exerts direct and indirect effects on IGF-I production. Estradiol (secreted or aromatized from testosterone) directly stimulates the production of IGF-I from cartilage and also increases GH secretion, which stimulates IGF-I production indirectly. Both actions appear important in the pubertal growth spurt.

D. Glucocorticoids   Endogenous or exogenous glucocorticoids in excess quickly decrease growth rate; this effect occurs more quickly than weight gain. The absence of glucocorticoids has little effect on growth if the individual is clinically well in other respects (eg, in the absence of hypotension or hypoglycemia).

Other Factors

A. Genetic factors   Genetic factors influence adult height. There is correlation between midparental height and the child's height; appropriate methods to utilize this phenomenon and determine the target height for a child are presented in Figure 6–6.
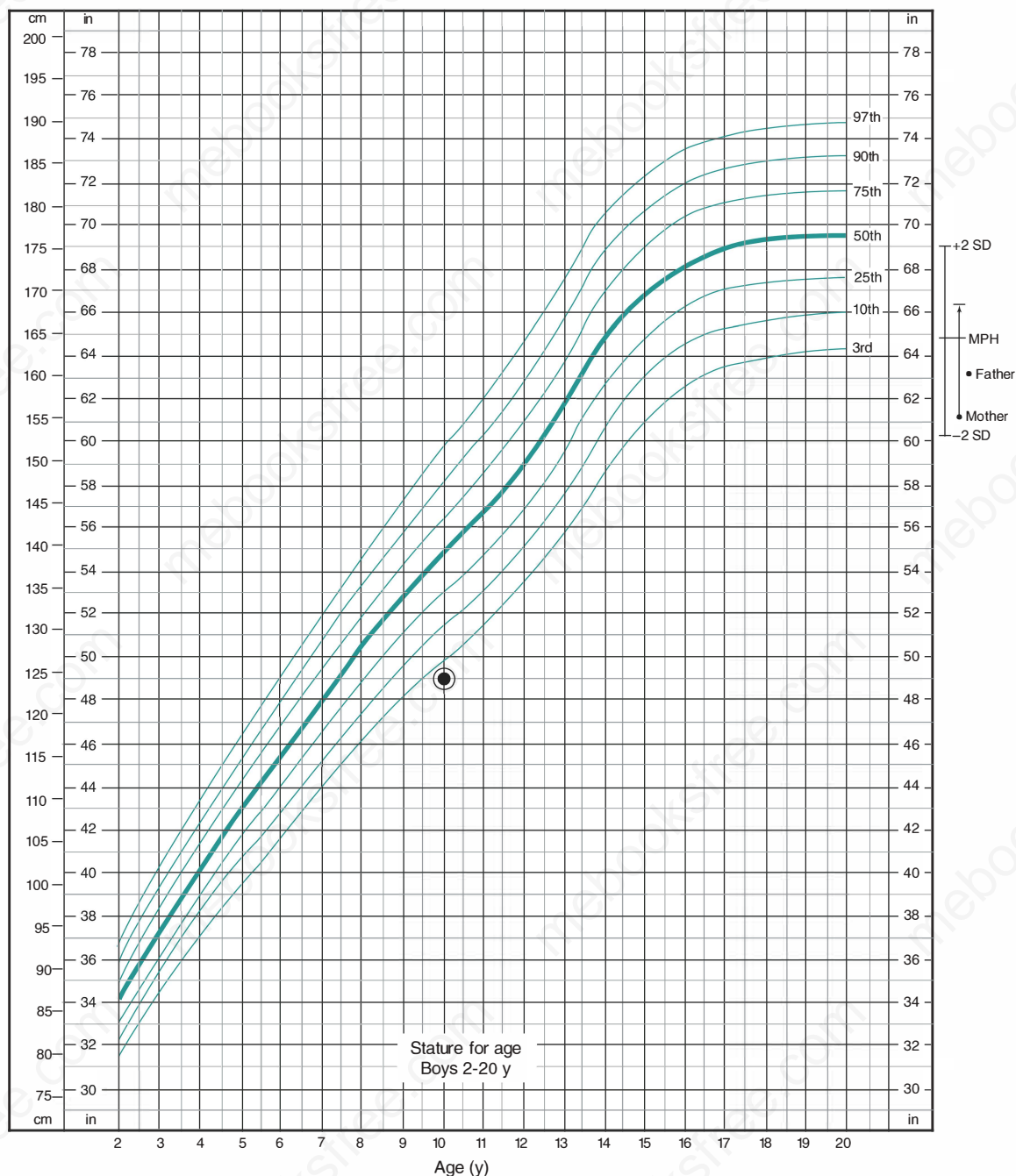
FIGURE 6–6 Determination of target height in a shorter family. This 10-year-old boy is 124 cm tall, his mother is 61 inches tall (adults recall their heights in inches and feet but, if they are available, their height should be actually measured in centimeters), and the father is 63 inches tall. Five inches are added to the mother's height to convert her height percentile to the equivalent percentile on a boy's chart. (If we were considering a daughter whose height is plotted on a girl's chart, the mother's height would be directly plotted and 5 inches would be subtracted from the father's height to correct his height percentile to the equivalent for an adult woman.) Her corrected height and that of the father are plotted at the far right of the chart where adult heights are displayed. The corrected midparental height is calculated by adding the father's height to the corrected mother's height, and the sum is divided by two; the result is called the target height. The limits of 2 SDs above and below the target height are displayed by plotting 2 SDs (~4 inches above and below the target height). This process is equivalent to moving the 50th percentile for the U.S. population to a conceptual 50th percentile for the family under consideration. It is evident that the height of the child, although below the 3rd percentile for the United States, is within the bounds of the percentiles described by ±2 SD from the target height, and the child appears to fit within the genetic pattern of the family. The growth velocity and the degree of skeletal maturation are some of the factors necessary to evaluate this child in more detail.

B. Socioeconomic factors Worldwide, the most common cause of short stature is poverty and its effects. Thus, poor nutrition, poor hygiene, and poor health influence growth both before and after birth. Parasitic infection is prevalent in less developed countries and severely stunts growth and depletes energy. In people of the same ethnic group and in the same geographic location, variations in stature are often attributable to socioeconomic factors. For example, Japanese individuals born and reared in North America after World War II were generally taller than Japanese-born immigrants moving to North America. Conversely, when other factors are equal, the differences in average height between various ethnic groups are mainly genetic. The newest growth charts for children in the United States published by the Centers for Disease Control and Prevention (CDC) in 2001 are not ethnicity-specific because it is believed that the major growth differences between ethnic groups are due to socioeconomic status and nutrition rather than genetic endowments. Indeed, the World Health Organization (WHO) has developed internationally relevant growth charts for all populations. The data for the construction of these charts is derived from well-nourished breastfed infants.

C. Nutritional factors The influence of malnutrition accounts for much of the socioeconomic discrepancy in height noted earlier, but malnutrition may occur in the midst of plenty and must always be suspected in disorders of growth. Other factors may be blamed for poor growth when nutritional deficiencies are actually responsible. For example, Sherpas were thought to have short stature mainly because of genetic factors or the effects of high altitude living on the slopes of Mount Everest; however, nutritional supplementation increased stature in this group, demonstrating the effects of adequate nutrition. The developed world places a premium on appearance, and women portrayed as beautiful in the media are characteristically thin. Significant numbers of children, chiefly teenagers, voluntarily decrease their caloric intake even if they are not obese; this accounts for some cases of poor growth. Chronic disease, which hampers adequate nutrition, often leads to short stature. For example, bronchopulmonary dysplasia decreases growth to some degree because it increases metabolic demands, shifting nutrient usage away from growth; improved nutrition increases growth in these patients. Celiac disease is another common gastrointestinal disorder that impairs growth, pubertal development, menstruation, and bone acquisition.

Feeding problems in infants, resulting from inexperience of parents or poor child–parent interactions (so-called maternal deprivation), may account for poor growth. Fad diets, such as poorly constructed vegan diets that put children at risk for vitamin B_{12} or iron deficiency, as well as major dietary manipulation, such as a severely low-fat diet, may place children at risk for deficiency of fat-soluble vitamins. Deliberate starvation of children by caregivers is an extreme form of child abuse that may be first discovered because of poor growth.

Remarkably, obesity increases IGF-I concentrations by increasing GH receptors even though GH secretion is suppressed to levels that might suggest GH deficiency, thereby complicating the diagnosis of true GH deficiency in the obese. There are significant endocrine changes associated with malnutrition. Decreased GH receptors or post-receptor defects in GH action, leading to decreased production of IGF-I and decreased concentration of serum IGF-I, are notable. The characteristic results of malnutrition are an elevation of serum GH, and a decrease in IGF-I. IGFBP-1, a suppressor of IGF-I effects, is elevated in malnutrition.

D. Psychologic factors Aberrant intrafamilial dynamics, psychologic stress, or psychiatric disease can inhibit growth either by altering endocrine function or by secondary effects on nutrition (psychosocial dwarfism or maternal deprivation). It is essential to diagnose those situations that might suggest organic disease states, as the management approach is very different in malnutrition.

E. Chronic disease Many chronic systemic diseases interfere with growth independent of poor nutrition. For example, congestive heart failure and asthma, if uncontrolled, are associated with decreased stature; in some cases, adult height is in the normal range because growth continues over a longer period of time. Children of mothers with HIV infection are often small at birth and have an increased incidence of poor postnatal growth, delayed bone age development, and reduced IGF-I concentrations. In addition, thyroid dysfunction may develop, further complicating the growth pattern. Infants born of HIV-infected mothers, who are not infected themselves, may exhibit catch-up growth.

Catch-up Growth

Correction of growth-retarding disorders may be temporarily followed by an abnormally high-growth rate as the child approaches normal height for age. This catch-up growth occurs after initiation of therapy for hypothyroidism and GH deficiency, after correction of glucocorticoid excess, and after appropriate treatment of many chronic diseases such as celiac disease. Catch-up growth is usually short-lived and is followed by a more typical growth rate. Catch-up growth after an SGA birth may not be beneficial as it is linked to metabolic disease, hypertension, obesity, and comorbid illnesses related to insulin resistance, by the age of 30 years.

MEASUREMENT OF GROWTH

Accurate measurement of height is an essential part of the physical examination of children and adolescents. The onset of a chronic disease may often be determined by an inflection point in the growth chart. In other cases, a detailed growth chart indicates a normal constant growth rate in a child observed to be short for age. If careful growth records are kept, a diagnosis of constitutional delay in growth and adolescence or genetic short stature may be made in such a patient; without previous measurements, the child might be subjected to unnecessary diagnostic testing, or months of delay may occur as the child's growth is finally carefully monitored. Poor measurement technique may suggest lack of

growth in a child who is growing normally, subjecting the child to unnecessary testing as well.

Height

The National Center for Health Statistics (NCHS) revised the growth charts for children in the United States in 2000 (see Figures 6–1 and 6–2). These charts display the 3rd and 97th percentiles rather than the 5th and 95th percentiles, and SDs of height for age are also available. Charts displaying BMI by age contain data appropriate for the evaluation of obesity and underweight. All of these charts are available online from the CDC.

The limits of the CDC growth charts showing the 3rd and the 97th percentile still leave 6 of 100 healthy children outside of their boundaries, with 3 of 100 below the more worrisome (to parents) *lower limits of normal*. It is both unnecessary and impractical to evaluate 3% of the population. Instead, the examining physician should determine which short children warrant further evaluation and which ones (and their parents) require only reassurance that they are healthy. When parents see that their child is below the 3rd percentile and in a section of the chart colored differently from the *normal area*, they assume that there is a serious problem. Thus, the format of the chart can dictate parental reaction to height, because all parents want their children to be in the *normal range*. Figures 6–1 and 6–2 furnish data necessary to evaluate the height of children at various ages using percentiles or the SD method used by the WHO. SD determination is more useful in extremely short children below the second or 1st percentile. CDC Z score data, equivalent to heights down to 2 SD below the mean for stature by age and gender is found at http://www.cdc.gov/growthcharts/zscore.htm.

Pathologic short stature is defined by different authorities in different manners. The Pediatric Endocrine Society (PES) uses less than 2 SD below the mean for age and gender or the 2.3 percentile, and the FDA uses –2.25 SDs or the 1.22 percentile for consideration of GH therapy for ISS. Since the CDC charts denote the 3% at the lowest line on the chart, one might approximate this to the (PES) definition of short stature. However, a diagnosis of pathologic short stature should not be based on a single measurement. Serial measurements are required because they allow determination of growth velocity, which is a more sensitive index of the growth process than a single determination. A very tall child who develops a postnatal growth problem will not fall below the 3rd percentile for years but will fall below the mean in growth velocity soon after the onset of the disorder. As Figures 6–3 and 6–4 demonstrate, growth velocity varies at different ages, but as a rough guide, a growth rate of less than 5 cm/y between age 4 years and the onset of puberty is abnormal. In children under 4 years of age, normal growth velocity changes more strikingly with age. A growth rate less than two SD below mean for age or height, corrected for midparental height, of less than –2 SD is worthy of attention. The Growth Hormone Research Society adds other criteria of concern: signs indicative of an intracranial lesion, signs of multiple pituitary hormone deficiencies, symptoms and signs of GH deficiency in the newborn and a history of CNS radiation. The more of the criteria met, the more likely is a disorder of growth.

Growth charts for premature infants ("Fenton growth curves") are available at http://pedinfo.org/growth.php. Growth charts for special populations are also available at that website; thus growth charts for achondroplasia, Down syndrome, Noonan syndrome, Williams syndrome, Turner syndrome, Russell-Silver syndrome, and other conditions can also be found.

Healthy term newborns tend to be clustered in length measurements around 21 inches (often owing to mistakes in obtaining accurate measurements). In the ensuing 24 postnatal months, the healthy child's height will enter a channel on the growth chart and remain there throughout childhood. Thus, a child with constitutional delay in growth or genetic short stature, whose height is at the mean at birth and gradually falls to the 10th percentile at 1 year of age and to the 5th percentile by 2 years of age, may in fact be healthy in spite of crossing percentile lines in the journey to a growth channel at the 5th percentile. Although the growth rate may decrease during these years, it should not be less than the 5th percentile for age. A steeper decrease in growth rate may be a sign of disease. Alternatively, catch-up growth after SGA birth will send the child to a higher height percentile. As stated earlier, catch-up growth from SGA is linked by some evidence to metabolic disease, particularly insulin resistance later in life. When a question of abnormal growth arises, previous measurements are essential. Every physician treating children should record supine length (<2 years of age) or standing height (>2 years of age) as well as weight at every office visit. As the child leaves infancy, height and growth velocity should be determined in relation to the standards for the child's age on a graphic chart with indication of the child's position (supine or standing) at measurement, which is especially important at the time children switch from lying to standing at about 2 years of age. Failure to recognize a change in measurement technique as the child moves from lying to standing may falsely suggest a growth problem.

Patients who cannot be measured in the standing position (eg, because of cerebral palsy) require other approaches. The use of arm span is a possible surrogate for the measurement of height, and there are formulas available for the calculation of height based on the measurement of upper arm length, tibial length, and knee length (see later).

This discussion presupposes accuracy of measurements. However, it is reported that screening examinations in the real world fall short of that ideal. Forty-one percent of a presumably normal population screened at a school in England met the criteria for evaluation of abnormal growth (approximately two-thirds grew faster than the normal growth category and one-third were in the slower than normal category), leading to an unreasonable size of a referral population, all due to simple measuring errors.

Relation to Midparental Height: The Target Height

There is a positive correlation between midparental height (the average of the heights of both parents) and the stature of a child. One way to use this relationship of parents' heights to the expected heights of children within a given family is to calculate target adult height range using parents' heights, correcting these

heights for the sex of the child and plotting the results on the child's growth chart. There is an average 5-inch height difference between adult men and women in the United States. Thus, for boys, add 5 inches to the mother's height, add the result to the father's height, and divide by two. This is the target height, and it is expected that sons of these parents will reach a height within two SD of this target—or, for simplicity, within 4 inches above or 4 inches below the target height (2 inches approximates one SD for adult stature). For girls, subtract 5 inches from the father's height and add the result to the mother's height and divide by two, leading to the target height for the girl. The range for girls will also be within 4 inches above and below this target. In effect, this corrects the North American growth charts for the particular family being considered. The calculated target height corresponds to the 50th percentile for the family, and the limits of the ±2 SD approximate the 5th to 95th percentile for the family. This method is useful only in the absence of disease affecting growth, and the prediction is more valid when the parents are of similar rather than of widely different heights. Figures 6–6 and 6–7 demonstrate the calculation of target height and the ranges. When there is a large discrepancy between the heights of the mother and the father, prediction of target height becomes difficult. A child may follow the growth pattern of the shorter parent more closely than the midparental height. A boy may, for example, follow the growth pattern of a short mother rather than a taller father.

A parent who spent the growing years in poverty, with chronic disease, or in an area of political unrest might have shorter adult height, due to nutritional factors or disease, that may not be passed on to the children. Of course, the height of an adopted child will have no relationship to the adoptive parents' heights. All of these factors must be determined by history.

Technique of Measurement

Length and height must be measured accurately. Hasty measurements derived from marks made on paper at an infant's foot and head while the infant is squirming on the paper on the examining table are useless. Infants must be measured on a firm horizontal surface with a permanently attached rule, a stationary plate perpendicular to the rule for the head, and a movable perpendicular plate for the feet. One person should hold the head stable while another makes sure the knees are straight and the feet are firm against the movable plate. There are calipers-like devices (eg, infantometer) that can be used for such accurate measurements.

Children over age two are measured standing up. Standing height is on the average 1.25 cm less than supine length, and it is essential to record the position of measurement each time during the change over from lying to standing height at 2 to 3 years of age; shifting from supine height at 2 years to standing height at 2½ years can falsely suggest an inadequate growth rate over that 6-month period.

Standing measurements cannot be accurately performed with the measuring rod that projects above the common weight scale; the rod is too flexible, and the scale footplate will in fact drop lower when the patient stands on it. Instead, height should be measured with the child standing back to the wall with heels at the wall, ankles together, and knees and spine straight against a vertical metal rule permanently attached to the wall or to a wide upright board. The child's head is to be horizontal with the eyes looking ahead and the chin gently elevated if necessary (Frankfort plane). Height is measured at the top of the head by a sliding perpendicular plate (or square wooden block). A Harpenden stadiometer is a mechanical measuring device capable of such accurate measurement. It is preferable to measure in the metric system, because the smaller gradations make measurements more accurate by minimizing the effect of rounding off numbers.

Growth is not constant but is characterized by short spurts and periods of slowed growth. The interval between growth measurements should be adequate to allow an accurate evaluation of growth velocity. Appropriate sampling intervals vary with age but should not be less than 3 months in childhood, with a 6-month interval being optimal.

The problem of measuring the growth rate of children with orthopedic deformities or contractures is significant, because these patients may have nutritional and/or endocrine disorders as well. The measurement of knee height, tibial length, or upper arm length correlates well with standing height ($r = 0.97$); thus, these measurements may be translated, using special linear regression equations, into total height, which is then plotted on standard growth charts. Specialized laser-calibrated devices to measure tibial length (kneeometry) are reported to be accurate for assessment of short-term growth down to weekly intervals.

In addition to height or length, other significant measurements include: (1) the frontal-occipital head circumference; (2) horizontal arm span (between the outspread middle fingertips with the patient standing against a flat backboard); and (3) the upper segment (US) to lower segment (LS) ratio. For the latter, the LS is measured from the top of the symphysis pubis vertically to the floor with the patient standing straight, and the US is determined by subtracting the LS from the accurate standing height measurement noted earlier. (Normal standard US-LS ratios are shown in Figure 6–5.) These measurements become important in skeletal or reproductive disorders. Sitting height is used in some clinical studies of growth, but the sitting stadiometer is rarely available.

Height and Growth Rate Summary

In summary, we consider three criteria for pathologic short stature: (1) height more than 2–2.25 SD below the mean for chronologic age; (2) growth rate more than 2 SD below the mean for chronologic age; and (3) height more than 2 SD below the target height when corrected for midparental height.

Weight and BMI

The measured weight should be plotted for age on standard graphs developed by the NCHS. These are available online (http://www.cdc.gov/growthcharts/clinical_charts.htm). BMI charts displaying percentiles of BMI (weight in kilograms divided by height in meters squared) are widely available and provide an essential method to assess nutritional status.
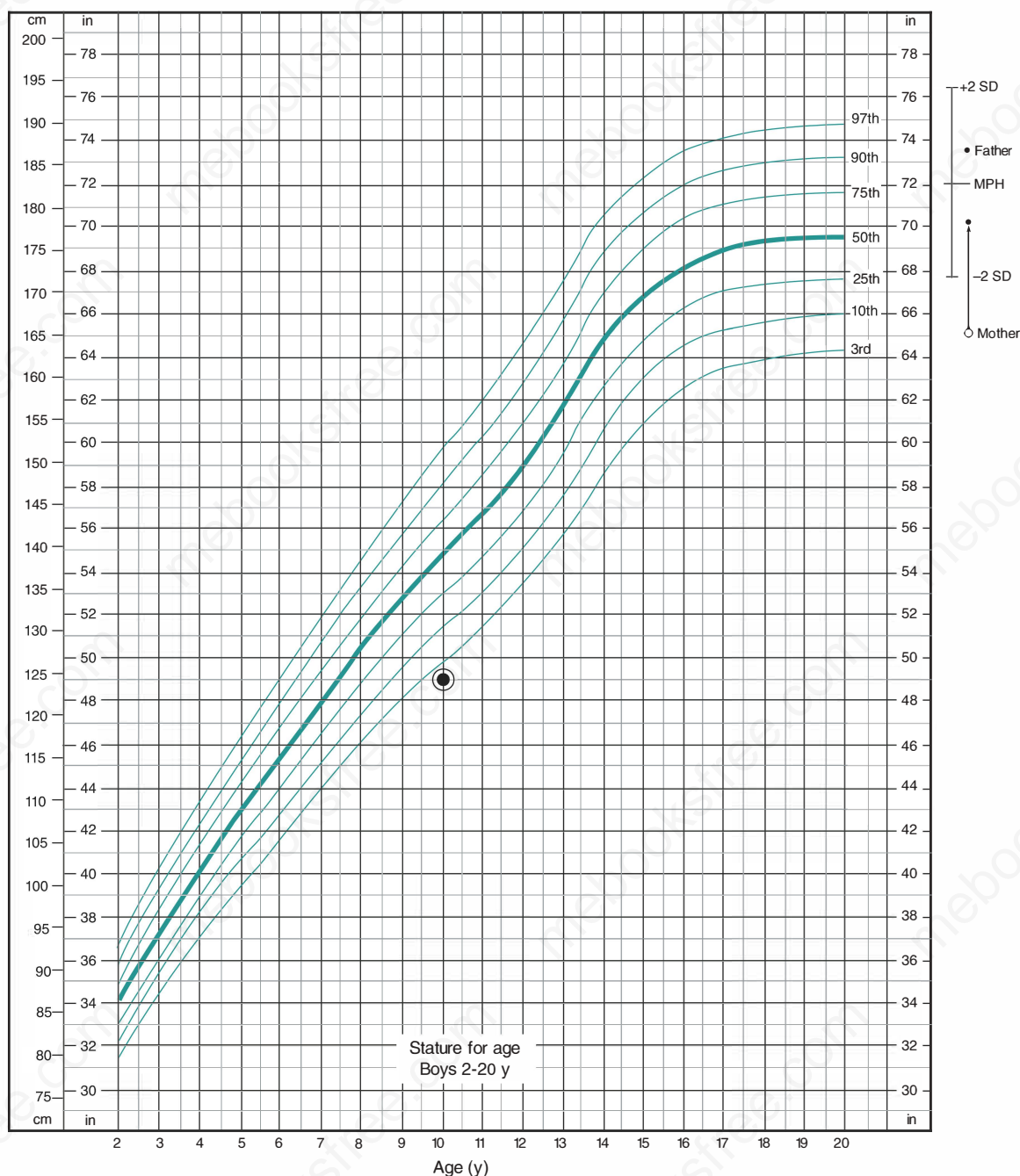
FIGURE 6–7 Determination of target height in a taller family. This 10-year-old boy is 124 cm tall just as in Figure 6–6, his mother is 65 inches tall, and the father is 73.5 inches tall. Five inches are added to the mother's height to convert her height percentile to the equivalent percentile on a boy's chart. (If we were considering a daughter whose height is plotted on a girl's chart, the mother's height would be directly plotted and 5 inches would be subtracted from the father's height to correct his height percentile to the equivalent height percentile for an adult woman.) Her corrected height and that of the father are plotted at the far right of the chart where adult heights are displayed. The corrected midparental height is calculated by adding the father's height to the corrected mother's height, and the sum is divided by 2; the result is the target height. The limits of 2 SDs above and below the target heights are displayed by plotting 2 SDs (about 4 inches above and below the target height). This is equivalent to moving the 50th percentile for the United States population to a conceptual 50th percentile for the family under consideration. It is evident that the height of the child, which is below the 3rd percentile for the United States, is even farther outside the bounds of the percentiles described by ±2 SD from the target height, and, thus, the child appears to fall far outside the genetic pattern of the family, and this is of clinical concern. The growth velocity and the degree of skeletal maturation are some of the other factors necessary to evaluate this child's growth in more detail.

SKELETAL (BONE) AGE

Skeletal development is a general reflection of physiologic maturation but there is significant variation. For example, menarche is better correlated with a bone age of 13 years than with a given chronologic age. However, bone age is as variable as chronologic age at the onset of puberty.

Estrogen plays the major role in advancing skeletal maturation. Patients with aromatase deficiency, who cannot convert testosterone to estradiol, and patients with estrogen receptor defects, who cannot respond to estrogen, grow taller well into their twenties without having epiphysial fusion. Bone age indicates remaining growth available to a child and can be used to predict adult height. However, bone age is not a definitive diagnostic test of any disease; it can assist in diagnosis only when considered along with other factors.

Bone age is determined by comparing the shapes and stage of fusion of epiphyses or bones on the patient's radiograph with an atlas demonstrating normal skeletal maturation for various ages. The Greulich and Pyle Atlas of radiographs of the left hand and wrist is most commonly used in the United States, but other methods of skeletal age determination, such as Tanner and Whitehouse maturity scoring, are preferred in Europe. Any bone age more than two SD above or below the mean for chronologic age is out of the normal range. The SD of bone age readings for chronological age is a full year by the mid teenage years, indicating the imprecision of prediction; thus at a chronological age of 15 years, the bone age can normally range from 13 to 17. Further, there appear to be ethnic differences in bone age maturation that is not reflected in the guidelines for interpretation. For newborn infants, knee and foot radiographs are compared with an appropriate bone age atlas. For late pubertal children, just before epiphyseal fusion, the knee atlas reveals whether any further growth can be expected or whether the epiphyses are fused. All the methods are imperfect, as there is great variation in bone age in relation to chronologic age even in typically developing children.

Height prediction in the United States utilizes the bone age reading and height measurement at the time the radiograph was taken and the Bayley-Pinneau tables in the Greulich and Pyle Skeletal Atlas of the hand to determine the amount of growth left before epiphyseal fusion. A chart to predict adult height of a typically developing child without the use of bone age is found in the Cole reference below. Height prediction by any method becomes more accurate as the child approaches the time of epiphyseal fusion.

DISORDERS OF GROWTH

SHORT STATURE DUE TO NONENDOCRINE CAUSES

There are many causes of decreased childhood growth and short adult height (Table 6–1). The following discussion covers only the more common conditions, emphasizing those that might be

TABLE 6–1 Causes of abnormalities of growth.

I. CAUSES OF SHORT STATURE

Nonendocrine Causes	Endocrine Disorders
Constitutional short stature	GH deficiency and variants
Genetic short stature	Congenital GH deficiency
Intrauterine growth retardation and SGA	Isolated GH deficiency
Syndromes of short stature	With other pituitary hormone deficiencies
Turner syndrome and its variants	With midline defects
Noonan syndrome	Pituitary agenesis
Bardet-Biedl syndromes	Holoprosencephaly
Other autosomal abnormalities and dysmorphic syndromes	Other anomalies
Chronic disease	Acquired GH deficiency
Cardiac disorders	Hypothalamic-pituitary tumors
Left to right shunt	Histiocytosis X
Congestive heart failure	Central nervous system infections
Pulmonary disorders	Head injuries
Cystic fibrosis	GH deficiency following cranial irradiation
Asthma	Central nervous system vascular accidents
Gastrointestinal disorders	Hydrocephalus
Malabsorption	Empty sella syndrome
Celiac disease	Abnormalities of GH action
Disorders of swallowing	Primary IGF-I deficiency
Hepatic disorders	Laron dwarfism
Hematologic disorders	Pygmies
Sickle cell anemia	IGF-I receptor defect
Thalassemia	Psychosocial dwarfism
Renal disorders	Hypothyroidism
Renal tubular acidosis	Glucocorticoid excess (Cushing syndrome)
Chronic uremia	Endogenous
Immunologic disorders	Exogenous
Connective tissue disease	Psuedohypoparathyroidism/ Albrights Hereditary Osteodystrophy
Juvenile rheumatoid arthritis	Disorders of vitamin D metabolism
Chronic infection	Diabetes mellitus, poorly controlled
Central nervous system disorders	
Malnutrition	
Decreased bioavailability of nutrients	
Fad diets	
Voluntary dieting	
Anorexia nervosa	
Anorexia of cancer chemotherapy	
Parasites	

II. CAUSES OF TALL STATURE

Nonendocrine Causes	Endocrine Disorders
Constitutional tall stature	Pituitary gigantism
Genetic tall stature	Sexual precocity
Syndromes of tall stature	Thyrotoxicosis
Cerebral gigantism	Infants of diabetic mothers
Marfan syndrome	
Homocystinuria	
Beckwith-Wiedemann syndrome	
XYY and XYYY syndromes	
Klinefelter syndrome	

included in an endocrine differential diagnosis. Shorter than average stature need not be considered a disease, because variation in stature is a normal feature of human beings, and a normal child should not be burdened with a misdiagnosis. Although the classifications described later may apply to most patients, some will still be resistant to definitive diagnosis.

1. CONSTITUTIONAL SHORT STATURE

Constitutional short stature (constitutional delay in growth and adolescence) is not a disease but rather a variation from normal for the population and is considered a slowing of the pace of development. There is usually an associated delay in pubertal development as well as a decrease in growth (see Constitutional Delay in Adolescence, Chapter 15). It is characterized by moderate short stature (usually not far below the 3rd percentile), thin habitus, and retardation of bone age. The family history often includes similarly affected members (eg, mother with delayed menarche or father who first shaved later than peers and continued to grow past his teen years).

All other causes of decreased growth must be considered and ruled out before this diagnosis can be made with confidence. The patient may be considered physiologically (but not mentally) delayed in development. Characteristic growth patterns include normal birth length and height, with a gradual decrease in percentiles of height for age by 2 years; in contrast, a rapid decrease in percentiles is an ominous sign of pathology. Onset of puberty is usually delayed for chronologic age but normal for skeletal age. Adult height is in the normal range but varies according to parental heights. Adult height is often less than the predicted height, because growth is less than expected during puberty. Aromatase inhibitors have been used in clinical studies for boys with constitutional delay in growth to inhibit the conversion of androgens to estrogen so that bone age does not advance and growth continues longer. While this is still an experimental treatment, most studies suggest that it causes no ill effects, although long-term data dealing with effects on bone mineral density as a result of the use of these agents are not available. Elevated testosterone levels and testicular enlargement result from aromatase inhibition.

2. FAMILIAL SHORT STATURE

Short stature may also occur in a familial pattern without retarded bone age or delay in puberty; this is considered **familial short stature**. Affected children are closer to the mean on the normal population growth charts after familial correction for mid-parental height by calculation of the target height (see Figures 6–6 and 6–7). Adult height depends on the mother's and father's heights. Patients with the combination of constitutional short stature and genetic short stature are quite noticeably short due to both factors, and these patients are most likely to seek evaluation. Boys are brought to consultation more often than girls. Children in these families may be born AGA but exhibit slowed growth in the first two postnatal years; this process is gradual in comparison to the striking changes in growth rate that are characteristic of diseases that primarily affect growth, but it may still be difficult to tell the difference without extended observation.

3. PREMATURITY AND SGA

Although the majority of SGA infants show catch-up growth, about 20% may follow a lifelong pattern of short stature. In comparison, AGA premature infants usually catch up to the normal range of height and weight by 1 to 2 years of age. Severely premature infants with birth weights less than 800 g (that are AGA), however, may maintain their growth retardation at least through their third year; only follow-up studies will determine whether this group of premature infants reaches reduced adult heights. Bone age and yearly growth rate are normal in SGA patients until puberty occurs, and the patients are characteristically thin. However, SGA is a risk factor for premature thelarche or early menarche, although catch up growth after SGA may be responsible for this tendency, so bone age may advance more rapidly in this subset of SGA (see Chapter 15).

Within this grouping are many distinctive genetic or sporadically occurring syndromes. The most common example is Russell-Silver dwarfism (OMIM #180860), characterized by small size at birth, triangular facies, and a variable degree of asymmetry of the extremities; this condition is due to epigenetic changes of DNA hypomethylation at the telomeric imprinting control region (ICR1) on chromosome 11p15, involving the *H19* and *IGF-II* genes or to maternal, uniparental disomy of chromosome 7. Intrauterine infections with *Toxoplasma gondii*, rubella virus, cytomegalovirus, herpesvirus, and HIV are noted to cause SGA. Furthermore, maternal drug usage, either illicit (eg, cocaine), legal but ill-advised (eg, alcohol during pregnancy), or legally prescribed medication (eg, phenytoin) may cause SGA. Reports of other syndromes in SGA infants can be found in sources listed in the bibliography.

Although SGA is not an endocrine cause of short stature, GH treatment is approved by the FDA and leads to increased adult height. Those SGA infants with the Δ3-isoform (genomic deletion of exon 3) of the GH receptor (GHR) are more likely to catch up on GH therapy, as are girls with Turner syndrome, although this benefit is transient and some data refute this differential response to treatment.

There are many endocrine sequelae with SGA birth including premature adrenarche, puberty, and menarche in girls and dyslipidemia and insulin resistance in boys and girls. Girls have a predilection to develop PCOS after being SGA. GH antagonizes the action of insulin and given the tendency for SGA children to have insulin resistance, there may be concerns regarding the potential additive effects. Recent studies of insulin sensitivity in SGA subjects receiving GH treatment indicate that in most cases the effects are not long lasting and do not seem to have clinical significance.

4. SYNDROMES OF SHORT STATURE

Many syndromes include short stature as a characteristic feature, and some also include SGA. Common conditions are described briefly later. Laurence-Moon, Biedl-Bardet, or Prader-Willi syndrome may combine obesity with short stature (as do the endocrine conditions, hypothyroidism, glucocorticoid excess, pseudohypoparathyroidism with Albright Hereditary Osteodystrophy (OMIM #103580), and GH deficiency). Moderately obese but otherwise normal children without these conditions tend to have slightly advanced bone age and advanced physiologic maturation with increased stature during childhood and early onset of puberty. Thus, short stature in an overweight child must be considered to have an organic cause until proven otherwise.

Turner Syndrome and Its Variants

Although classic Turner syndrome of 45, XO gonadal dysgenesis (see Chapter 14) is often correctly diagnosed, it is not always appreciated that any phenotypic female with short stature may have a variant of Turner syndrome. Thus, a karyotype determination should be performed for every short girl if no other cause for short stature is found, especially if puberty is delayed (see Chapter 15). The short stature of Turner syndrome is due to a mutation of the *short stature homeobox (SHOX)* gene on the short (p) arm in the pseudoautosomal region of the X chromosome (OMIM #312865). A mutation of the *SHOX* gene may also cause the Léri-Weill dyschondrosteosis form of short-limbed dwarfism (OMIM #127300).

Noonan Syndrome (Pseudo-Turner Syndrome) (OMIM #163950)

This syndrome shares several phenotypic characteristics of Turner syndrome, including short stature, webbed neck, low posterior hairline, and facial resemblance to Turner syndrome, but the karyotype is 46,XX in the female or 46,XY in the male with Noonan syndrome, and other features clearly differentiate it from Turner syndrome. For example, in Turner syndrome, there is characteristically left-sided heart disease and in Noonan syndrome right-sided heart disease. Noonan syndrome is an autosomal dominant disorder at gene locus 12q24 (see Chapter 14). GH therapy is approved by the FDA for Noonan syndrome to increase height. About half of patients with Noonan syndrome have a mutation of the protein tyrosine phosphatase nonreceptor type 11 (PTPN11) (OMIM #176876). These children have a decreased response to GH treatment and tend to have low IGF-I and ALS with normal IGFBP-3 levels. Noonan syndrome patients are prone to develop neoplasia and the PTPN11 mutation, among others, increases the risk adding a concern over GH treatment in them and underscores the need for surveillance for cancer.

Prader-Willi Syndrome (OMIM #176270)

This condition is characterized by poor intrauterine movement, acromicria (small hands and feet), developmental delay, and almond-shaped eyes along with infantile hypotonia. Short stature is common but not invariable. Although hypotonia limits feeding in infancy, later insatiable hunger develops and leads to extreme obesity. Glucose intolerance and delayed puberty are characteristic. This syndrome is due to deletion of the small nuclear riboprotein polypeptide N (SNRPN) on paternal chromosome 15 (q11-13), uniparental disomy of maternal chromosome 15, or methylation of this region of chromosome 15 of paternal origin. If a mutation of the same locus is derived from the mother, Angelman syndrome results. GH therapy is approved by the FDA for Prader-Willi syndrome to increase height, but the most important effects are improvement of body composition and muscle strength. There are several reported cases of the development of obstructive sleep apnea with GH treatment, which can be fatal, and so sleep patterns must be carefully monitored, and a sleep study must be performed, if there is question of sleep apnea before the administration of GH is begun.

Bardet-Biedl Syndrome

These autosomal recessive, but genetically heterogeneous conditions, are characterized by obesity, short stature, mental retardation, kidney dysfunction, polydactyly, and retinitis pigmentosa. Hypogonadotropic hypogonadism and primary hypogonadism have variously been reported in affected patients (#209900). There are presently 19 identified subtypes of this condition listed in Online Mendelian Inheritance in Man.

Autosomal Chromosome Disorders and Syndromes

Numerous other autosomal disorders and syndromes of dysmorphic children with or without developmental disorders are characterized by short stature. Often the key to diagnosis is the presence of several major or minor physical abnormalities that indicate the need for karyotype determination. Other abnormalities may include unusual body proportions, such as short extremities, leading to aberrant US:LS ratios, and arm spans quite discrepant from stature. Details of these syndromes can be found in the references listed at the end of the chapter.

Skeletal Dysplasias

There are more than 100 known types of genetic skeletal dysplasias (osteochondrodysplasias). Often they are noted at birth because of the presence of short limbs or trunk, but some are only diagnosed after a period of postnatal growth. The most common condition is autosomal dominant achondroplasia (OMIM #100800). This condition is characterized by short extremities in the proximal regions, a relatively large head with a prominent forehead due to frontal bossing and a depressed nasal bridge, and lumbar lordosis in later life. Adult height is decreased, with a mean of 132 cm for males and 123 cm for females. Intelligence is normal. Mutations of the tyrosine kinase domain of the *fibroblast growth factor receptor 3 (FGFR3)* gene locus 4p16.3 (OMIM #134934) have been identified in this condition. Limb lengthening operations are used to increase stature in a few centers but the techniques are complex, and complications appear to be frequent.

Height, height velocity, weight, and BMI curves from birth to 16 years of age are available for achondroplasia at http://pedinfo.org/growth.php. Children with achondroplasia who have received GH have in some instances demonstrated improved growth; however, atlanto-occipital dislocation is reported. The potential for abnormal brain growth and its relationship to aberrant skull shape mandates the caution that GH is not established therapy for this condition.

Hypochondroplasia (OMIM #146000) is manifested on a continuum from severe short-limbed dwarfism to apparent normal development until puberty, when there is an attenuated or absent pubertal growth spurt, leading to short adult stature. This disorder may be caused by an abnormal allele in the same gene causing achondroplasia (*FGFR3* at locus 4p16.3).

5. CHRONIC DISEASE

Severe chronic disease involving any organ system can cause poor growth in childhood and adolescence. In many cases, there are adequate historical or physical findings by the time of consultation to permit diagnosis. In some cases, however—most notably celiac disease and regional enteritis—short stature and decreased growth may precede obvious signs of malnutrition or gastrointestinal disease. In some cases, growth is only delayed and may spontaneously improve. In others, growth can be increased by improved nutrition; patients with gastrointestinal disease, kidney disease, or cancer may benefit from nocturnal parenteral nutritional infusions.

Cystic fibrosis combines several causes of growth failure: lung disease impairs oxygenation and predisposes to chronic infections, gastrointestinal disease decreases nutrient availability, and late-developing abnormalities of the endocrine pancreas cause diabetes mellitus. Children with cystic fibrosis experience decreased growth rates after 1 year of age following a normal birth size. The pubertal growth spurt is often decreased in magnitude and delayed in its timing; secondary sexual development may be delayed, especially in those with impaired pulmonary function. Study of growth in these patients allowed development of a cystic fibrosis specific growth chart and indicates that a reasonable outcome is an adult height in the 25th percentile. GH treatment in several studies demonstrated improved pulmonary function and increased growth and weight gain in cystic fibrosis. While growth charts for subjects with cystic fibrosis are available, use of standard charts for monitoring is also suggested.

Children with congestive heart failure, due to a variety of congenital heart diseases or acquired myocarditis, grow poorly unless successfully treated with medications or surgery. Patients with cyanotic heart disease and pulmonary hypertension appear to be most affected.

Celiac disease is present in about 1% of the population and may present initially with growth failure, but delayed puberty and menarche, osteopenia, osteoporosis, and other findings can arise. Early diagnosis can be made by determination of tissue transglutaminase antibodies, if IgA values are in the normal range while on a normal wheat-containing diet. However, a biopsy may still be required for diagnosis. On a gluten-free diet, patients experience catch-up growth. Adult height may still be impaired, depending on the duration of the period without treatment. Untreated patients with celiac disease have decreased serum IGF-I concentrations, presumably due to malnutrition, while IGF-I concentrations rise with dietary therapy. Thus, serum IGF-I in this condition, as in many with nutritional deficiencies, is not a reliable indicator of GH secretory status. Crohn's disease is associated with poor growth in 15% to 40% of cases. There is poor quality of evidence over the best manner to treat the condition and increase growth rate.

Patients with chronic hematologic diseases, such as sickle cell anemia or thalassemia, often have poor growth, delayed puberty, and short adult stature: iron deposition may itself cause endocrine complications. Juvenile rheumatoid arthritis may compromise growth before or after therapy with glucocorticoids. GH treatment is reported to increase the growth rate of these children. Hypophosphatemic vitamin D-resistant rickets usually leads to short adult stature, but treatment with 1,25-hydroxyvitamin D and oral phosphate in most cases will improve bone growth as well as increase adult stature. Chronic renal disease is known to interfere with growth, but increased growth rate occurs with improved nutrition. GH therapy is approved for affected children who still grow poorly.

Proximal or distal renal tubular acidosis (RTA) may cause short stature. Proximal RTA demonstrates bicarbonate wasting at normal or low plasma bicarbonate concentrations; patients have hypokalemia, alkaline urine pH, severe bicarbonaturia, and later, acidemia. The condition may be inherited, sporadic, or secondary to many metabolic or medication-induced disorders. Distal RTA is caused by inability to acidify the urine; it may occur in sporadic or familial patterns or be acquired as a result of metabolic disorders or medication. Distal RTA is characterized by hypokalemia, hypercalciuria, and occasional hypocalcemia. The administration of bicarbonate is the primary therapy for proximal RTA, and proper treatment can substantially improve growth rate.

Obstructive sleep apnea is associated with poor growth. The amount of energy expended during sleep in children with sleep apnea appears to limit weight and length gain, a pattern which reverses with the resolution of the obstruction. Obstructive sleep apnea is also associated with obesity and Prader-Willi syndrome. Diagnosis is made by polysomnography.

6. MALNUTRITION

Malnutrition (other than that associated with chronic disease) is the most common cause of short stature worldwide and is the cause of much short stature in the developing world. Diagnosis in the developed world is based on historical and physical findings, particularly the dietary history. Food faddism and anorexia nervosa—as well as excessive voluntary dieting—can cause poor growth. Infections with parasites, such as *Ascaris lumbricoides* or *Giardia lamblia*, can decrease growth and are a principal cause of short stature in the developing world.

Specific nutritional deficiencies can have particular effects on growth. For example, severe iron deficiency can cause a thin habitus as well as growth retardation. Zinc deficiency can cause anorexia, decreased growth, delayed puberty, developmental problems, hair loss, diarrhea, impotence, eye and skin abnormalities, and taste changes leading to weight loss. Acrodermatitis enteropathica is a rash giving a visual clue to Zn deficiency. It arises usually in the presence of chronic systemic disease or infection. Children with nutritional deficiencies characteristically demonstrate failure of weight gain before growth rate decreases, and before weight for height decreases. This is in contrast to many endocrine causes of poor growth, where weight for height remains in the normal or high range. This simple rule often determines whether nutritional or endocrine evaluations are most appropriate. There are no simple laboratory tests for diagnosis of malnutrition. Serum IGF-I concentrations are low in malnutrition, as they are in GH deficiency. This distinction is important since misdiagnosing GH deficiency instead of malnutrition would be tragic as well as costly.

7. MEDICATIONS

Children with hyperactivity disorders (or those incorrectly diagnosed as such) are frequently managed with chronic methylphenidate administration or similar medication. These agents can decrease weight gain by decreasing appetite and can lower growth rate, albeit inconsistently. These drugs must be used in moderation and only in children who definitely respond to them during careful evaluation and follow-up.

Exogenous glucocorticoids are a potent cause of poor growth (discussed later), as are excessive endogenous glucocorticoids.

SHORT STATURE DUE TO ENDOCRINE DISORDERS

Using the conservative criteria of height (<3rd percentile) and growth velocity (<5 cm/y for inclusion in the study), the incidence of endocrine disease in 114,881 Utah children who met these cutoff results was 5%, with a higher incidence in boys than girls by a ratio of more than 2.5:1. In this population, 48% of the children with Turner syndrome or GH deficiency were not diagnosed prior to the careful evaluation afforded by this study.

1. GROWTH HORMONE DEFICIENCY AND ITS VARIANTS

The incidence of GH deficiency is estimated to be between 1:4000 and 1:3500, so the disorder should not be considered rare.

There may be abnormalities at various levels of the hypothalamic-pituitary, GH-IGF-I axis. Most patients with idiopathic GH deficiency lack GHRH. One autopsied GH-deficient patient had an adequate number of pituitary somatotrophs that contained considerable GH stores. Thus, the pituitary gland produced GH, but it could not be released. Long-term treatment of such patients with GHRH can cause GH release and increase growth, but this therapy is not presently in use. Patients with pituitary tumors or those rare patients with congenital absence of the pituitary gland lack somatotrophs. Several kindreds have been described that lack various regions of the *GH* gene (responsible for producing GH). Alternatively, gene defects responsible for the embryogenesis of the pituitary gland may cause multiple pituitary deficiencies. Absence of the *PIT1* gene encoding a pituitary-specific transcription factor causes deficient GH, TSH, and prolactin synthesis and secretion. Mutations of the prophet of PIT1 (*PROP1*) gene cause deficiencies of GH, TSH, FSH, LH, and ACTH production.

Congenital Growth Hormone Deficiency

Congenital GH deficiency presents with slightly decreased birth length (−1 SD), but the growth rate decreases in some cases strikingly soon after birth. The disorder is identified by careful measurement in the first year and becomes more obvious by 1 to 2 years of age. Patients with classic GH deficiency have short stature, increased fat mass leading to a chubby or cherubic appearance with immature facial appearance, immature high-pitched voice, and delay in skeletal maturation. Less severe forms of partial

GH deficiency are described with few abnormal characteristics apart from short stature, decreased growth rate, and delayed bone age. GH-deficient patients lack the lipolytic effects of GH, partially accounting for the pudgy appearance. There is a higher incidence of hyperlipidemia with elevated total cholesterol and low-density lipoprotein (LDL) cholesterol in GH deficiency, and longitudinal studies demonstrate increases in high-density lipoprotein (HDL) cholesterol levels with GH treatment. Males with GH deficiency may have microphallus (penis <2 cm in length at birth), especially if the condition is accompanied by gonadotropin-releasing hormone (GnRH) deficiency (Figure 6–8). GH

FIGURE 6–8 A 12-month-old boy with congenital hypopituitarism. He had hypoglycemic seizures at 12 hours of age. At 1 year, he had another hypoglycemic seizure (plasma glucose, 25 mg/dL) associated with an episode of otitis media, and it was noted that his penis was extremely small. At 12 months, body length was 66.5 cm (−2 SD) and weight was 8.5 kg (−3 SD). The penis was less than 1.5 cm long, and both testes were descended (each 1 cm in diameter). Plasma GH did not rise above 1 ng/mL after arginine and levodopa testing (no insulin tolerance test was performed because of the history of hypoglycemia). LH rose very little after administration of GnRH (gonadorelin), 100 μg. Serum thyroxine was low (T_4, 6.6 μg/dL; T_4 index, 1.5), and after administration of 200 μg of protirelin (TRH), serum TSH rose with a delayed peak characteristic of tertiary hypothyroidism. Plasma ACTH rose only to 53 pg/mL after metyrapone. Thus, the patient had multiple defects in the hypothalamic-pituitary axis including decreased secretion of GH, ACTH, and TSH due to deficient secretion of hypothalamic hormones. He was given six doses of 2000 units each of human chorionic gonadotropin (hCG) intramuscularly over 2 weeks, and plasma testosterone rose to 62 ng/dL, indicating normal testicular function. He was then treated with 25 mg of testosterone enanthate every month for 3 months, and his phallus enlarged to 3.5 × 1.2 cm without significant advancement of bone age. With hGH therapy (0.05 mg/kg intramuscularly every other day), he grew at a greater than normal rate for 12 months (catch-up growth), and growth then continued at a normal rate.

deficiency in the neonate or child can also lead to symptomatic hypoglycemia and seizures; if ACTH deficiency is also present, hypoglycemia is usually more severe. The differential diagnosis of neonatal hypoglycemia in a full-term infant who has not sustained birth trauma must include neonatal hypopituitarism. If microphallus (in a male subject), optic hypoplasia, or some other midline facial or CNS defect is noted, the diagnosis of congenital GH deficiency is more likely (see later). Congenital GH deficiency is also correlated with breech delivery. Intelligence is normal in GH deficiency unless repeated or severe hypoglycemia is present or a significant anatomic defect has compromised brain development. When thyrotropin-releasing hormone (TRH) deficiency is also present, there may be additional signs of hypothyroidism. Secondary or tertiary congenital hypothyroidism is not usually associated with physical findings of cretinism or developmental delay as is congenital primary hypothyroidism, but a few cases of isolated TRH deficiency and severe mental retardation are reported.

Congenital GH deficiency may present with midline anatomic defects. Optic hypoplasia with visual defects ranging from nystagmus to blindness is found with variable hypothalamic endocrinopathies in 71.7% of one series: 64.1% of subjects had GH axis abnormalities, 48.5% hyperprolactinemia, 34.9% hypothyroidism, 17.1% adrenal insufficiency, and 4.3% diabetes insipidus (DI) in this group of 47 subjects. About half of patients with optic hypoplasia have absence of the septum pellucidum on computed tomography (CT) or magnetic resonance imaging (MRI), leading to the diagnosis of septo-optic dysplasia. Septo-optic dysplasia is most often sporadic in occurrence but some affected individuals are reported with mutations of the *homeobox* gene expressed in ES cells (*HESX1*) (OMIM #601802) and septo-optic dysplasia (OMIM #182230). Cleft palate or other forms of oral dysraphism are associated with GH deficiency in about 7% of cases. Such children may need nutritional support to improve their growth. An unusual midline defect associated with GH deficiency is described in children with a single maxillary incisor.

Congenital absence of the pituitary gland, which occurs in an autosomal recessive pattern, leads to severe hypopituitarism with hypoglycemia. Affected patients have shallow development of or absence of the sella turcica. This defect is quite rare but clinically devastating due to ACTH deficiency if treatment is delayed. This is the most common MRI manifestation of *PROP1* gene mutation (OMIM #601538).

Hereditary GH deficiency is described in several different mutations. Various genetic defects of the *GHN* gene (17q22-24) occur in affected families. Isolated type 1A GH deficiency (IGH-DIA OMIM #262400) is inherited in an autosomal recessive pattern. Patients have deletions, frameshifts, and nonsense mutations in the *GH* gene. Unlike those with classic sporadic GH deficiency, some of these children are reported with significantly shortened birth lengths. Patients with absent or abnormal *GH* genes initially respond to exogenous human GH (hGH) administration, but some soon develop high antibody titers that eliminate the effect of therapy. Patients with high titers of blocking antibodies are reported to benefit from IGF-I therapy in place of GH therapy. Isolated GH deficiency (IGHD) type 1B (OMIM #612781) patients have autosomal recessive splice site mutations

and incomplete GH deficiency and are less severely affected. Type 2 (IGHD2) (OMIM #173100) patients have autosomal dominant GH deficiency due to splice site or missense mutations. Type 2 patients have X-linked GH deficiency often associated with hypogammaglobulinemia.

A few patients are described with abnormalities of the *GHRH* gene. Mutations in pituitary transcription factors can lead to various combinations of pituitary hormone deficiencies as noted earlier. Mutations in *PIT1/POU1F1* (POU class 1 homeobox 1) (OMIM #173110) lead to GH, PRL, and TSH deficiencies. Mutations in *PROP1* (Prophet of PIT1, paired-like homeodomain transcription factor) (OMIM #601538) lead to GH, PRL, TSH, LH, FSH, and sometimes ACTH deficiencies. Mutations in *HESX1* lead to GH, PRL, TSH, LH, FSH, ACTH, IGHD, and CPHD deficiencies.

Acquired Growth Hormone Deficiency

Onset of GH deficiency in late childhood or adolescence, particularly if accompanied by other pituitary hormone deficiencies, is ominous and may be due to a hypothalamic-pituitary tumor. The development of posterior pituitary deficiency, in addition to anterior pituitary deficiency, makes a tumor even more likely. The empty sella syndrome is more frequently associated with hypothalamic-pituitary abnormalities in childhood than in adulthood; thus, GH deficiency may be found in affected patients.

Some patients, chiefly boys with constitutional delay in growth and adolescence, may have transient GH deficiency on testing before the onset of puberty. When serum testosterone concentrations begin to increase in these patients, GH secretion and growth rate also increase. This transient state may incorrectly suggest *bona fide* GH deficiency but does not require therapy. A priming dose of estrogen is sometimes invoked to increase GH secretion maximally so that spurious GH deficiency is not diagnosed spuriously. CNS conditions that cause acquired GH deficiency (eg, craniopharyngiomas, germinomas, gliomas, histiocytosis X) are described in Chapter 4. It is remarkable that after craniopharyngioma removal, some patients, mainly obese subjects, continue to grow quite well in spite of the absence of GH secretion. This persistent growth appears to be caused by hyperinsulinemia.

Cranial irradiation of the hypothalamic-pituitary region to treat CNS tumors or acute lymphoblastic leukemia may result in GH deficiency starting approximately 12 to 18 months later, owing to radiation-induced hypothalamic (or perhaps pituitary) damage. Higher doses of irradiation, such as the 24-Gy regimen previously used for the treatment of CNS leukemia, have greater effect (adult height may be as much as 1.7 SD below the mean) than the lower (eg, 18 Gy) regimens and higher doses are more likely to cause TSH, ACTH, and gonadotropin deficiency as well as hyperprolactinemia or even precocious puberty. Girls treated at an early age with this lower regimen still appear to be at risk of growth failure. All children must be carefully observed for growth failure after irradiation since growth failure may occur years later. If these patients receive spinal irradiation, upper body growth may also be impaired, causing a decreased US-LS ratio and a further

tendency to short stature. Abdominal irradiation for Wilms tumor may also lead to decreased spinal growth (estimated loss of 10 cm height from megavoltage therapy at 1 year of age, and 7 cm from treatment at 5 years of age). Others receiving gonadal irradiation (or chemotherapy) have impaired gonadal function, lack onset or progression of puberty, and/or have diminished or absent pubertal growth spurt. Chemotherapy for acute lymphocytic leukemia without irradiation may also lead to GH deficiency, so follow-up of growth after treatment for cancer is always necessary.

CNS trauma is well established as a cause of hypopituitarism in adults. Cross sectional studies in children reveal both high and low rates of subsequent hypopituitarism after head trauma, but the rare prospective studies demonstrate lower risk in children than in adults.

Other Types of GH Dysfunction

Primary IGF-I deficiency is due to GH insensitivity (Laron syndrome and its variants (OMIM #262500)). These disorders reflect GH receptor or post-receptor defects that are inherited in autosomal recessive fashion. The soluble GH-binding protein (GHBP) found in the circulation arises from the extracellular portion of the GH receptor, and since they derive from the same gene, circulating GHBP reflects the abundance of GH receptors. Patients with decreased or absent GH receptors have decreased serum GHBP levels, whereas those with post-receptor defects have normal GHBP concentrations. Affected children are found throughout the world, including Israel, where the syndrome was first reported, and Ecuador, where several generations of a large kindred were studied in great detail. Defects in various kindreds include nonsense mutations, deletions, RNA processing defects, and translational stop codons. GH/JAK-STAT axis signal-transduction impairment leads to short stature, when this intracellular system fails to activate in response to receptor occupancy by the GH ligand. Defects in the dimerization of the GH receptors, a required step in GH action, also lead to short stature. Serum GH is elevated in all forms of GH resistance, due to decreased or absent IGF-I, which results in lack of negative feedback inhibition. Patients are short at birth, confirming the importance of IGF-I in fetal growth (demonstrated previously in murine *IGF-I* gene knockout studies). About one-third have hypoglycemia, and half of boys have microphallus. The condition does not respond to GH treatment. Patients treated with recombinant IGF-I grow at an improved rate but do not respond quite as well to IGF-I as GH-deficient children do to GH treatment, indicating that GH may have a direct role in fostering growth above and beyond that conferred by IGF-I.

Other forms of GH resistance are described, but the majority of patients with disorders of the GH axis have abnormalities of GH secretion, not action. Very short, poorly growing children with delayed skeletal maturation, normal GH and IGF-I values, and no signs of organic disease have responded to GH therapy with increased growth rates equal to those of patients with bona fide GH deficiency. These patients may have a variation of constitutional delay in growth or genetic short stature, but a subtle abnormality of GH secretion or action is possible.

Pygmies (OMIM #265850) have normal serum GH, low IGF-I after puberty, and normal IGF-II concentrations. They have a congenital inability to raise IGF-I concentrations after puberty, which has greater importance in stimulating growth than IGF-II. Pygmy children are reported to lack a pubertal growth spurt, suggesting that IGF-I is essential to attain a normal peak growth velocity. Efe pygmies, the shortest of the pygmies, are significantly smaller at birth than neighboring Africans who are not pygmies, and their growth is slower throughout childhood, leading to statures displaced progressively below the mean. A few patients are reported with defects of the *IGF-I* gene or with deficiency of the IGF-I receptor (IGF-IR) (OMIM #147370) and extreme short stature that is not responsive to IGF-I. Intrauterine growth deficiency, microcephaly, developmental delay, and other psychological findings are reported in these cases. Why do certain normal children, perhaps within a short family, have stature significantly lower than the mean? There is no definite answer to this persistent question, but some patients have decreased serum GHBP concentration, which suggests a decrease in GH receptors in these children. A minority of short, poorly growing children have definable genetic abnormalities of their GH receptors. Genome Wide Association studies are uncovering various genetic influences on normal and pathologic stature. Short stature is the final common pathway of numerous biochemical abnormalities.

Adults who had GH deficiency in childhood or adolescence have decreased bone mass compared with normals even when bone mass is corrected for their smaller size. There is progressive bone loss in adults who are GH deficient even if bone density was improved with childhood therapy; GH is approved for adults with GH deficiency and can reverse this trend (see Chapter 4).

Diagnosis of GH Deficiency

Because basal values of serum GH are low in normal children and GH-deficient patients alike, the diagnosis of GH deficiency has classically rested on demonstration of an inadequate rise of serum GH after provocative stimuli or on some other measure of GH secretion. This process is complicated because different radioimmunoassay systems vary widely in their measurements of GH in the same blood sample (eg, a result on a single sample may be above 10 ng/mL in one assay but only 6 ng/mL in another). The physician must be familiar with the standards of the laboratory being used. Most insurance companies and state agencies accept inability of GH to rise above 10 ng/mL with stimulation as diagnostic of GH deficiency.

Another complicating factor is the state of pubertal development. Prepubertal children secrete less GH than pubertal subjects and, especially as they approach the onset of puberty, may have sufficiently reduced GH secretion to suggest falsely *bona fide* GH deficiency. This factor is sometimes addressed by administering a dose of estrogen to such subjects before testing. The very concept of GH testing provides a further complication. GH is released in episodic pulses. Although a patient who does not secrete GH in response to standard challenges is generally considered to be classically GH-deficient, a normal GH response to these tests may not rule out the efficacy of GH treatment. Testing should occur after

an overnight fast; carbohydrate or fat ingestion suppresses the GH response. Obesity suppresses GH secretion, and an overweight or obese child may falsely appear to have GH deficiency. Even within the normal range, variations of BMI affect peak GH after stimulation.

Because 10% or more of healthy children do not have an adequate rise in GH with one test of GH reserve, at least two methods of assessing GH reserve are necessary before the diagnosis of classic GH deficiency is assigned. Of course, if GH rises above 10 ng/mL in a single test, classic GH deficiency is eliminated. Serum GH values should rise after 10 minutes of vigorous exercise; this is used as a screening test. After an overnight fast, GH levels should rise in response to arginine infusion (0.5 g/kg body weight [up to 20 g] over 30 minutes), oral levodopa (125 mg for up to 15 kg body weight, 250 mg for up to 35 kg, or 500 mg for >35 kg), or clonidine (0.1 mg/m^2 orally). Side effects of levodopa include nausea; those of clonidine include some drop in blood pressure and drowsiness. Glucagon stimulation testing is used to determine both GH and ACTH secretory ability. It is accomplished by administration of 30 μg/kg glucagon (maximum 1 mg) and collecting GH samples at 0, 30, 60, 90, 120, 150, and 180 minutes afterwards; nausea and hyperglycemia are possible side effects.

The insulin tolerance test is another way to assess GH reserve but can be dangerous to perform and is rarely invoked. GH levels rise after acute hypoglycemia due to insulin administration; however, this test carries a risk of seizure if the blood glucose level drops excessively. An insulin tolerance test may be performed if a 10% to 25% dextrose infusion is available for emergency administration in the face of hypoglycemic coma or seizure and if the following conditions are satisfied: (1) an intravenous infusion line with heparin lock or low-rate saline infusion is available before the test begins, (2) the patient can be continuously observed by a physician, and (3) the patient has no history of hypoglycemic seizures. The patient must have a normal glucose concentration at the beginning of the test in the morning after an overnight fast (water intake is acceptable). Regular insulin, 0.075 to 0.1 U/kg in saline, may be given as an intravenous bolus. In 20 to 40 minutes, a 50% drop in blood glucose will occur, and a rise in serum GH and cortisol and ACTH should follow. Serum glucose should be monitored, and an intravenous line must be maintained for emergency dextrose infusion in case the patient becomes unconscious or has a hypoglycemic seizure. If dextrose infusion is necessary, it is imperative that blood glucose not be raised far above the normal range, because hyperosmolality has been reported from overzealous glucose replacement; undiluted 50% dextrose should not be used (see Chapter 4). This is a dangerous test and is rarely indicated!

A family of synthetic penta- and hexapeptides called GH-releasing peptides (GHRPs) stimulate GH secretion in normal individuals and in GH-deficient subjects. GHRPs act via ghrelin receptors of the hypothalamus that are different from the GH-releasing factor (GHRF) receptors, and their effects are additive to that of GHRF. Ghrelin is a gastrointestinal-derived peptide that naturally binds to these receptors in the hypothalamic ventromedial nucleus, arcuate nucleus and ventral tegmental area (dopaminergic neurons), causing GH secretion, it also stimulates appetite and is classified as an orexigenic agent. Recently, mutations in the

GH secretagogue receptor (GHSR) that binds GHRPs and ghrelin were found in children with short stature; treatment with GH increased growth rate in these children.

Patients who respond to the pharmacologic stimuli noted earlier, but not to physiologic stimuli, such as exercise or sleep, are said to have neurosecretory dysfunction. These patients have decreased 24-hour secretion of GH (or integrated concentrations of GH) compared with healthy subjects, patterns similar to those observed in GH-deficient patients. It is not clear how frequently this condition is encountered.

This long discussion of the interpretation of GH after secretagogue testing brings into question the very standard for the diagnosis of GH deficiency. It is clear that pharmacologic testing cannot always determine which patients truly need GH therapy, and many authorities suggest that we abandon such dynamic testing in favor of measurements of IGF-I and IGFBP-3, although dynamic GH testing may still be required by insurance plans. Serum IGF-I values are low in most GH-deficient subjects, but, as noted earlier, some short patients with normal serum IGF-I concentrations may require GH treatment to improve growth rate. Moreover, starvation lowers IGF-I values in healthy children and incorrectly suggests GH deficiency. Children with psychosocial dwarfism—who need family therapy or foster home placement rather than GH therapy—have low GH and IGF-I concentrations and may falsely appear to have GH deficiency. Likewise, patients with constitutional delay in adolescence have low IGF-I values for chronologic age, but normal values for skeletal age, and may have temporarily decreased GH response to secretagogues. Thus, IGF-I determinations are not infallible in the diagnosis of GH deficiency. They must be interpreted with regard to nutrition, psychosocial status, and skeletal ages. IGFBP-3 is GH-dependent, and if its concentration is also low, it provides stronger evidence of GH deficiency than does the IGF-I determination alone.

The Growth Hormone Research Society produced criteria that attempt to deal with the diagnosis of GH deficiency in childhood in spite of the uncertainty of the methods. These criteria use clinical findings of various conditions associated with GH deficiency, the severity of short stature, and the degree and duration of decreased growth velocity to identify individuals that may have GH deficiency. The guidelines and diagnostic considerations in this chapter include most of the GH Society criteria. (See "Consensus guidelines" reference in the Short Stature section at the end of this chapter for details of the Growth Hormone Research Society statement.) A 3- to 6-month therapeutic trial of GH therapy may be necessary to determine growth response; if growth increases more than 2 cm/y, it is likely that the child will benefit from GH treatment, no matter what the tests originally showed.

Treatment of GH Deficiency

A. GH replacement   Before 1986, the only available method of treatment for GH deficiency was replacement therapy with hGH derived from cadaver donors. In 1985 and thereafter, Creutzfeldt-Jakob disease, a degenerative neurologic disease rare in patients so young, was diagnosed in some patients who had received natural hGH up to 10 to 15 years before. Because of

the possibility that prions contaminating donor pituitary glands were transmitted to the GH-deficient patients, causing their deaths, natural GH from all sources was removed from distribution. Recombinant hGH now accounts for the world's current supply.

Commercial hGH has the 191-amino acid natural sequence. hGH is now available in virtually unlimited amounts, allowing innovative treatment regimens not previously possible owing to scarce supplies. The potential for abuse of hGH in athletes or in children of normal size whose parents wish them to be taller than average, however, must now be addressed.

Growth disturbances due to disorders of GH release or action are shown in Table 6–2. GH-deficient children require biosynthetic somatropin (natural sequence hGH) at a dose of 0.18–0.3 mg/kg/wk administered in one subcutaneous dose per day 6 or 7 times per week during the period of active growth before epiphyseal fusion. An increased dose is approved by the FDA for use during puberty. The increase in growth rate (Figures 6–9, 6–10, and 6–11) is most marked during the first year of therapy. Older children do not respond as well and may require larger doses. Higher doses, up to double the standard starting dose, are approved by the FDA for use in puberty, but there are varying reports of the effect of these higher doses on adult height as, by far, most of the effect of GH on adult height is exerted in the years before puberty. GH does not increase growth rate without adequate nutrition and euthyroid status. During the roughly 50 years since the first use of GH in children, long-term effects are reported in several series. If only the children most recently treated with recombinant hGH are considered, the mean adult height was 1.4 SD below the mean, a significant improvement over the –2.9 SD mean height at the start of therapy but not a true normalization of height in most patients. With earlier diagnosis and treatment, using new pubertal dosing schedules, adult height can reach genetic potential.

TABLE 6–2 Level of defect in growth.

Site of Defect	Clinical Condition
Hypothalamus	Idiopathic GH deficiency due to decreased GHRH secretion; hypothalamic tumors or congenital defects
Pituitary gland	Dysplasia, trauma, surgery, or tumor of the pituitary gland; defect in GH gene or in pituitary transcription factors
Sites of IGF production	Primary IGF-I deficiency: GH receptor defect (Laron dwarfism with high GH and low IGF concentrations) ALS deficiency Growth hormone/JAK-STAT axis signal transduction defect Pygmies with normal GH, low IGF-I, and normal IGF-II concentrations
Cartilage	Glucocorticoid-induced growth failure Resistance to IGF-I

Monitoring of GH replacement is mainly accomplished by measuring growth rate and annually assessing bone age advancement. Serum IGF-I and IGFBP-3 will rise with successful therapy while GHBP will not change appreciably. Controlled clinical studies reported the utility of titrating the dose of GH to restore serum IGF-I to the high-normal range, monitoring serum IGF-I levels during clinical treatment. There is concern if the IGF-I levels rise more than two SDs above the mean and, if so, the GH dose is decreased. Serum bone-specific alkaline phosphatase rises with successful therapy. Urinary hydroxyproline, deoxypyridinoline, and galactosyl-hydroxylysine reflect growth rate and are used in clinical studies to reflect increased growth rate with therapy.

Antibodies to GH may be present in measurable quantities in the serum of children receiving GH. However, a high titer of blocking antibodies with significant binding capacity is rare except in patients with absence or abnormality of GH genes. Only a few patients are reported to have temporarily ceased growing because of antibody formation. GH exerts anti-insulin effects. Although clinical diabetes is not a likely result of GH therapy, the long-term effects of a small rise in glucose in an otherwise healthy child are unknown. If a tendency toward diabetes is already present, GH may cause the onset of clinical manifestations more quickly. Another potential risk is the rare tendency to develop slipped capital femoral epiphyses in children receiving GH therapy; slipped capital femoral epiphyses occur at times of increased growth rate. Recent data have weakened the link between slipped capital femoral epiphyses and GH therapy, but the final import of the relationship is not yet clear. Slipped capital femoral epiphyses, if associated with endocrinopathies, are most common in treated hypothyroid patients (50% one series of 80 episodes of slipped capital femoral epiphyses), followed by treated GH-deficient patients (25% of the series). This condition may occur bilaterally, and prophylactic treatment of the nonaffected side is recommended by several authorities. Pseudotumor cerebri may rarely occur with GH therapy, is usually associated with severe headache and may be more common in obese individuals receiving GH treatment. It is reported to reverse after cessation of GH therapy, but if allowed to continue, it may impair vision due to pressure on the optic nerve and cause severe complications. Organomegaly and skeletal changes like those found in acromegaly are other theoretical side effects of excessive GH therapy but do not occur with standard doses. Furthermore, prepubertal gynecomastia is reported with GH therapy.

Lack of compliance with GH therapy is a frequent cause of poor growth leading to undertreatment. Reported non-compliance is in the range of 33%. Lack of normalization of IGF-I with treatment is an indication of noncompliance.

The discovery of leukemia in young adults previously treated with GH was worrisome, but no cause and effect relationship has been established, and GH treatment is not considered a cause of leukemia. GH does not increase the recurrence rate of tumors existing before therapy. Thus, patients with craniopharyngiomas, for example, may receive GH, if indicated, after the disease is clinically stable, without significant worry that the GH will precipitate a recurrence. Clinicians usually wait 1 year after
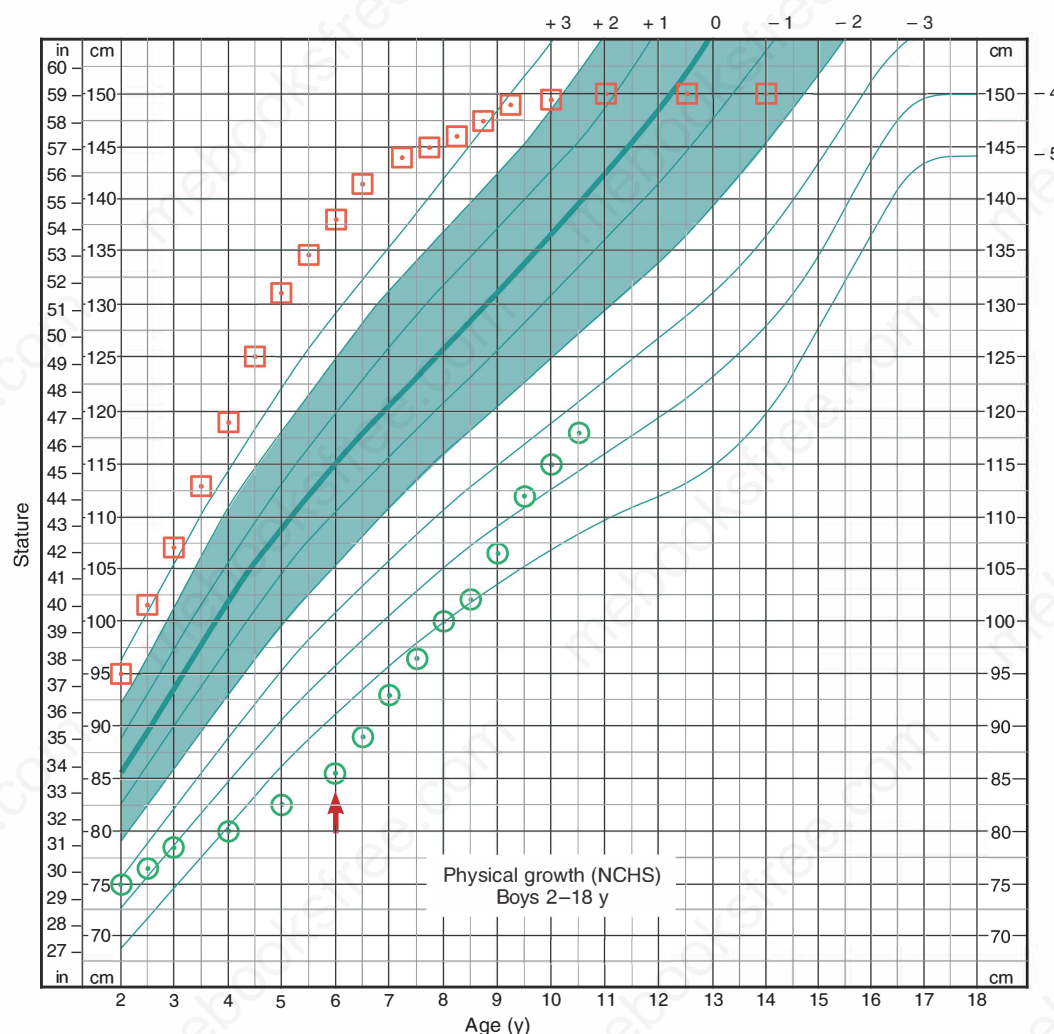
FIGURE 6–9 Examples of abnormal growth charts. Squares represent the growth pattern of a child (such as patient **A** in Figure 6–11) with precocious sexual development and early excessive growth leading to premature closure of the epiphyses and cessation of growth. Circles represent growth of a boy (such as patient **B** in Figure 6–11) with GH deficiency who showed progressively poorer growth until 6 years of age, when he was treated with hGH (arrow), after which catch-up growth occurred. The curves describe SDs from the mean.

completion of tumor therapy before starting patients on GH therapy, but doing so is not a requirement. There are reports of a small increase in risk of colonic carcinoma decades after natural GH treatment in GH-deficient children, but no such information on long-term follow-up of children treated with recombinant hGH is available. GH deficiency is associated with an adverse lipid profile with elevated LDL cholesterol and decreased HDL cholesterol in addition to an increased BMI; GH-deficient adolescents treated with GH develop these findings within a few years after discontinuation of GH therapy. Low-dose GH therapy is now approved for use in adults with childhood-onset GH deficiency and is said to forestall these metabolic changes. Further, adult GH therapy maintains muscle strength and bone density in GH-deficient adults. One can, therefore, inform the parents of a child with GH deficiency that the patient may still benefit from GH therapy even after he or she stops growing if profound GH deficiency remains after repeat testing once off GH for at least 1 month.

GH has been combined with other substances to increase its impact on height. In patients who were diagnosed late, have entered puberty, and appear to have limited time to respond to GH before epiphyseal fusion causes the cessation of growth, a GnRH agonist has been used to delay epiphyseal fusion in clinical trials with varying success, but this was not recommended by a consensus conference on the use of GnRH agonist therapy, due to lack of strong evidence of effectiveness. This off-label use is not yet established as safe and effective. Aromatase inhibitors have been combined with GH in some clinical studies, but this is not yet regular clinical practice; there was a reported effect of decreasing bone age advancement, while allowing increased growth and height.

There are other conditions for which the FDA has approved the use of GH. GH therapy will increase adult height in Turner syndrome to an average of 5.1 cm if started early enough; the addition of low-dose oxandrolone is reported to further increase
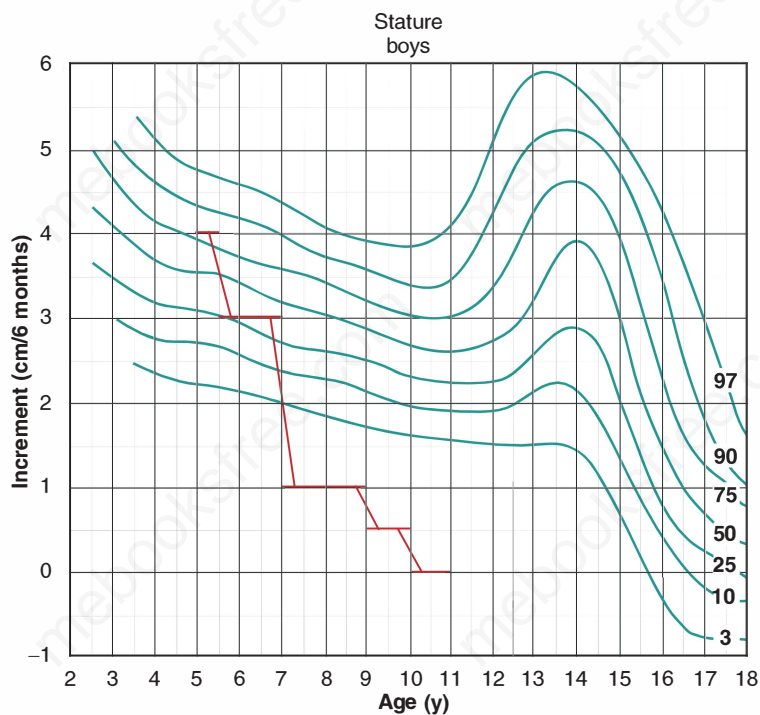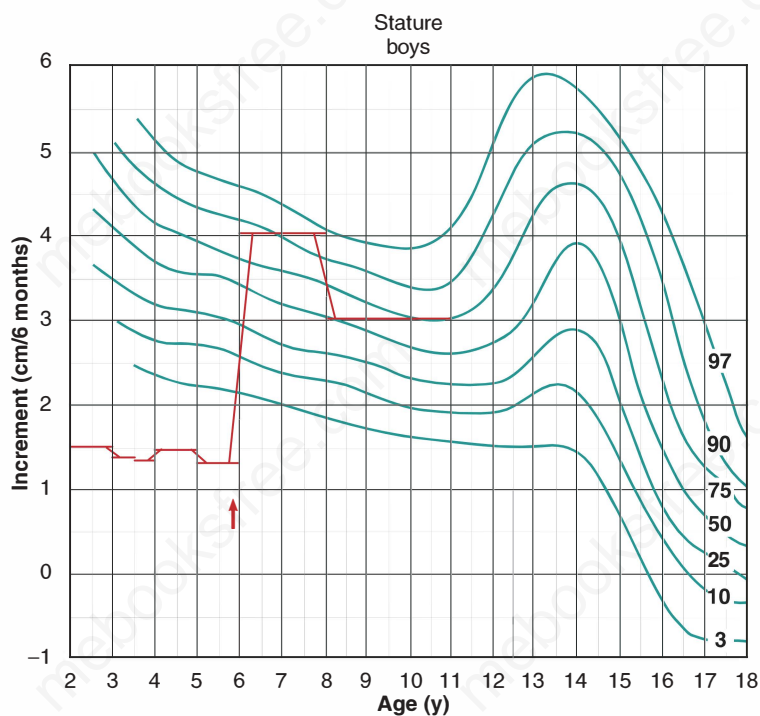
FIGURE 6–10 Two examples of abnormal growth plotted on a height velocity chart simplified from the charts in Figure 6–3. **A**. The plot is taken from the data recorded as squares in Figure 6–9, describing a patient with precocious puberty such as patient **A** in Figure 6–11, with premature epiphysial closure, and cessation of growth. **B**. The plot is taken from the data recorded as circles in Figure 6–9, describing a patient with GH deficiency (such as patient **B** in Figure 6–11) who was treated with hGH (arrow) at age 6. Initial catch-up growth is noted for 2 years, with a lower (but normal) velocity of growth following. These charts display growth rate over 6 month growth intervals rather than 12-month intervals as shown on Figures 6–3 and 6–4.
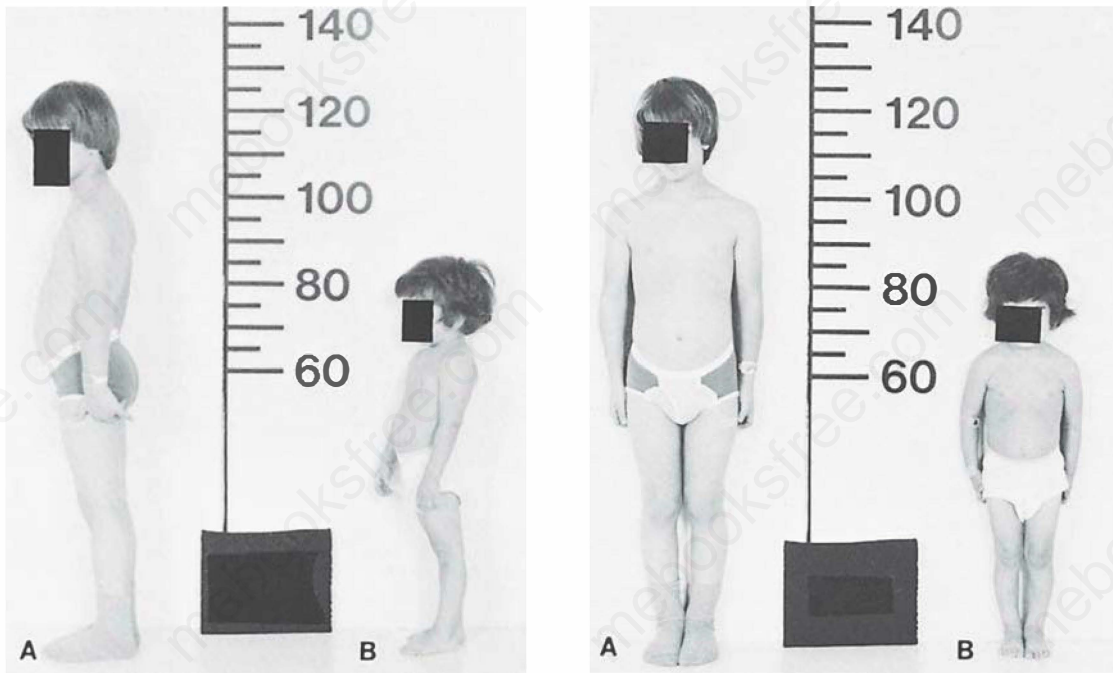
FIGURE 6–11 Two boys demonstrating extremes of growth. The boy at left in each photograph **(A)** has precocious puberty due to a CNS lesion. At 4½ years, he was 125.1 cm tall, which is 4.5 SDs above the mean. (The mean height for a 4-year-old is 101.5 cm.) His testes measured 2 × 3.5 cm each, his penis 9.8 × 2.8 cm (all pubertal measurements). He was muscular and had acne and a deep voice. His bone age was 10 years, the testosterone level was 480 ng/dL, and the LH rose after 100 μg of GnRH (gonadorelin) to a pubertal response. His brain CT scan revealed a hamartoma of the tuber cinereum. The boy at right **(B)** at 6 years of age was 85 cm tall, which is more than 5 SD below the mean. He had the classic physical and historical characteristics of idiopathic GH deficiency, including early growth failure and a cherubic appearance. His plasma GH values were nondetectable and did not rise after provocative testing.

growth rate. Estrogen is used to promote feminization and maintain bone mineral density; the optimal time to initiate treatment with estrogen should be individualized based on bone age, height, and psychological factors. Usually estrogen is administered during the adolescent years in low doses and only after the normal age of onset of puberty is reached to preserve maximal adult height, although earlier initiation of therapy is gaining credence (see Chapter 15).

Recently, French researchers released data from a large ongoing prospective study and reported an increased risk of hemorrhagic stroke in early adulthood after hGH therapy during childhood and adolescence. However the cause and effect of hGH treatment on stroke remain to be established. The FDA is presently investigating these data, but findings have not changed clinical practice in the United States, and the FDA has not suggested that the clinical use of GH change. Present evidence does not support an increase in all cause mortality or neoplasia from hGH treatment.

In the most successful series of children with SGA treated with hGH, the agent increased adult height between 2.0 SD and 2.7 SD. Girls with Turner syndrome treated with hGH reach adult height of more than 150 cm, an increase from the average untreated height which is about 144 cm. When treatment starts at or before 4 years, an adult height is achieved in the normal range.

Treatment with GH is approved for chronic renal disease in childhood. GH increases growth rate above the untreated state without excessive advancement in bone age. Prader-Willi syndrome may also be treated with GH to increase growth rate, lean tissue mass, and bone density. A recent study indicated that parental education exerted the most significant effect on body composition in these children. However, there are reports that patients with Prader-Willi syndrome have died of obstructive sleep apnea following GH treatment, demonstrating a need for sleep studies to exclude sleep apnea before initiating treatment and constant surveillance after treatment has been initiated. The treatment of Noonan syndrome is described earlier. In general, males are more often treated with GH than females in the United States, but this is not the case in other developed countries.

The FDA has approved the use of GH in otherwise normal children, whose stature is below 2.25 SDs for age and who are predicted to fall short of reaching normal adult height (<1 percentile of adult height). There may be pressure for treatment of children predicted to be taller than these guidelines from parents, but the FDA approval is for specific indications. While GH may increase the height of such severely affected children, it should not be used for a child whose predicted adult height is in the normal range. Treatment costs $30,000 to $40,000 per year or about $35,000 per cm gained. In the United States, males are treated more often than females, while the ratio is more equal in other countries, presumably due to social/cultural issues.

GHRH has been isolated, sequenced, and synthesized. It is available for use in diagnosis and treatment. GH-deficient patients demonstrate lower or absent GH secretion after administration of

GHRH. However, episodic doses of GHRH can restore GH secretion, IGF-I production, and growth in children with idiopathic GH deficiency. The ability of GHRH administration to stimulate pituitary GH secretion further supports the concept that idiopathic GH deficiency is primarily a disease of the hypothalamus, not of the pituitary gland.

IGF-I is now produced by recombinant DNA technology. IGF-I is useful in the treatment of certain types of short stature, particularly Laron dwarfism (and perhaps for African pygmies, should treatment be desired) where neither GH nor any other treatment is effective. IGF-I has been studied clinically for more than 12 years. Side effects, such as hypoglycemia (observed in 49% of treated subjects), injection site lipohypertrophy (32%), and tonsillar/adenoidal hypertrophy (22%), are common but not said to be severe. However, there is concern among some authorities as to the safety of this therapy.

B. Psychologic management and outcome   Research into the psychologic outcome of patients with short stature is flawed by lack of consistent methods of investigation and lack of controlled studies, but some results are of interest.

Studies vary in their conclusions, as to whether short stature is harmful to a child's psychologic development or not, and whether, by inference, GH is helpful in improving the child's psychologic functioning. Children with GH deficiency are the most extensively studied; earlier investigations suggested that they have more passive personality traits than do healthy children, may have delayed emotional maturity, and suffer from infantilization from parents, teachers, and peers. Many of these children have been held back in school because of their size without regard to their academic abilities. Some patients retain a body image of short stature even after normal height has been achieved with treatment. More recent studies challenge these views and suggest that self-image in children with height below the 5th percentile, who do not have GH deficiency, is closely comparable to a population of children with normal height. These findings may not be representative of the patient population discussed earlier in that a normal ambulatory population of *short* children may differ from the selected group that seeks medical attention. The data suggest that short stature itself is not cause for grave psychologic concern, and such concerns should not be used to justify GH therapy. We cannot avoid the fact that our *heightist* society values physical stature and equates it with the potential for success, a perception that is not lost on the children with short stature and their parents. A supportive environment in which they are not allowed to act younger than their age, nor to occupy a privileged place in the family is recommended for children with short stature. Psychologic help is indicated in severe cases of depression or maladjustment.

2. PSYCHOSOCIAL DWARFISM (FIGURE 6–12)

Children with psychosocial dwarfism present with poor growth, a *pot-bellied* and immature appearance. They often display bizarre eating and drinking habits. Parents may report that the affected child begs for food from neighbors, forages in garbage cans, and drinks from toilet bowls. As a rule, this tragic condition occurs in only one of several children in a family. Careful questioning and observation reveal a disordered family structure in which the child is either ignored or severely disciplined. Caloric deprivation or physical battering may or may not be a feature of the history. These children have functional hypopituitarism. Testing often reveals GH deficiency at first, but after the child is removed from the home, GH function quickly returns to normal. Diagnosis rests on improvement in behavior or catch-up growth in the hospital or in a foster home. Separation from the family is therapeutic, but the prognosis is guarded. Family psychotherapy may be beneficial, but long-term follow-up is lacking.

Growth disorder due to abnormal parent–child interaction in a younger infant is *maternal deprivation* no matter which parent is most closely associated with the condition. Caloric deprivation due to parental neglect may be of greater significance in this younger age group. Even in the absence of nutritional restriction or full-blown psychosocial dwarfism, constant negative interactions within a family may inhibit the growth of a child.

It is essential to consider family dynamics in the evaluation of a child with poor growth. It is not appropriate to recommend GH therapy for emotional disorders.

3. HYPOTHYROIDISM

Thyroid hormone deficiency decreases postnatal growth rate and skeletal development. Congenital hypothyroidism leads to severe developmental delay unless treatment is rapidly provided after birth. Screening programs for the diagnosis of congenital hypothyroidism have been instituted all over the world. Early treatment following diagnosis in the neonatal period markedly reduces growth failure and has virtually eliminated developmental abnormalities caused by this disorder. Early treatment of congenital hypothyroidism results in normal growth. Acquired hypothyroidism in older children (eg, due to lymphocytic thyroiditis) may lead to growth failure. Characteristics of hypothyroidism are decreased growth rate and short stature, retarded bone age, and an increased US-LS ratio for chronologic age due to poor growth of the extremities. Patients are apathetic and sluggish and have constipation, bradycardia, coarsening of features and loss of hair, hoarseness, and delayed pubertal development, if the condition is untreated. Intelligence is unaffected in late-onset hypothyroidism, but the apathy and lethargy may make it seem otherwise. Although weight gain is possible with hypothyroidism, in contrast to common wisdom, it is not extreme.

The diagnosis of congenital hypothyroidism is usually made on the basis of neonatal screening studies. In this standard procedure, currently in use throughout the world, a sample of blood is taken from the heel or from the umbilical cord at birth and analyzed for total T_4 or TSH. A low total T_4 or free T_4 for age or an elevated TSH is usually indicative of congenital hypothyroidism; actual values differ by state and laboratory. A low total T_4 alone may be associated with low circulating thyroxine-binding globulin (TBG), but the significantly elevated TSH is diagnostic of primary hypothyroidism. The diagnosis may be accompanied by radiologic
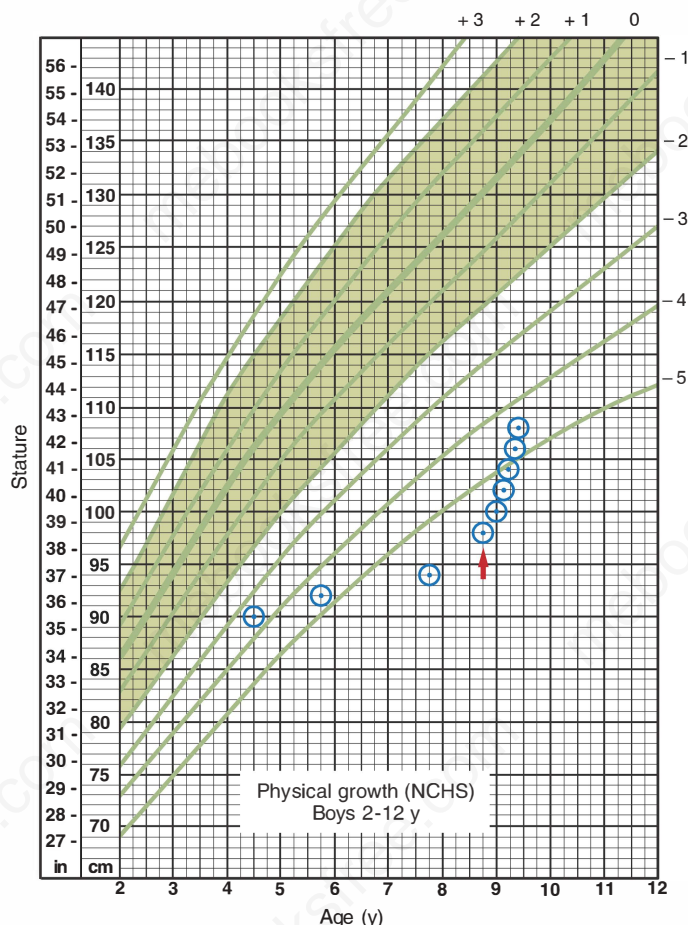
FIGURE 6–12   Photograph and growth chart of a 9½-year-old boy with psychosocial dwarfism. He had a long history of poor growth (<3 cm/y). The social history revealed that he was given less attention and punished more frequently than his seven siblings. He ate from garbage cans and begged for food, though he was not apparently deprived of food at home. When the photograph was taken, he was 99 cm tall (–7 SD) and weighed 14.7 kg (–3 SD). His bone age was 5 years, with growth arrest lines visible. Serum thyroxine was 7.8 µg/dL. Peak serum GH varied from nondetectable to 8 ng/mL on different provocative tests between age 6 years and 8½ years. He was placed in a hospital chronic care facility (arrow) for a 6-month period and grew 9 cm, which projects to a yearly growth velocity of 18 cm. On repeat testing, the peak serum GH was 28 ng/mL.

evidence of retarded bone age or epiphyseal dysgenesis in severe congenital hypothyroidism.

In older children, serum TSH is the most reliable diagnostic test. Elevated TSH with decreased free T_4 may eliminate the potential confusion resulting from the use of total T_4, which may vary with the level of TBG or other thyroxine-binding proteins. A positive test for serum thyroglobulin antibodies or thyroperoxidase antibodies would lead to the diagnosis of autoimmune thyroid disease (Hashimoto thyroiditis) as an explanation for the development of hypothyroidism (see Chapter 7). If both FT_4 and TSH are low, the possibility of central hypothyroidism (pituitary or hypothalamic insufficiency) must be considered (this would not be revealed in a newborn screening programs that use primarily TSH screening). This should lead to a search for other hypothalamic-pituitary endocrine deficiencies such as GH deficiency and CNS disease (see Chapter 4).

Treatment is accomplished by thyroxine replacement. The dose varies from a range of 10 to 15 µg/kg in infancy to 2 to 3 µg/kg

in older children and teenagers. Suppression of TSH to normal values for age is a useful method for assessing the adequacy of replacement in acquired primary hypothyroidism. However, there are additional considerations in treatment of neonates as suppression of TSH is not appropriate in all affected newborns, and consultation with a pediatric endocrinologist is essential in this age group to ensure optimal dosing and adequate CNS development.

4. CUSHING SYNDROME

Excess glucocorticoids (either exogenous or endogenous) lead to decreased growth before obesity and other signs of Cushing syndrome develop. The underlying disease may be bilateral adrenal hyperplasia due to abnormal ACTH-cortisol regulation in Cushing disease, autonomous adrenal adenomas, or adrenal carcinoma. The appropriate diagnosis may be missed if urinary cortisol and 17-hydroxycorticosteroid determinations are not interpreted on the

basis of the child's body size or if inappropriate doses of dexamethasone are used for suppressive testing (appropriate doses are 20 μg/kg/d for the low-dose and 80 μg/kg/d for the high-dose dexamethasone suppression test) (see Chapter 9). Furthermore, daily variations in cortisol production necessitate several urinary or plasma cortisol determinations before Cushing disease can be appropriately diagnosed or ruled out. The high-dose dexamethasone test was positive in 68% of a recent series of children with Cushing disease. The corticotropin-releasing hormone test was positive in 80% of affected patients, whereas MRI of the pituitary was positive in only 52%. Inferior petrosal sampling (see Chapters 4 and 9) was 100% accurate in the diagnosis of Cushing disease, although technically difficult in children. The development of salivary cortisol assays makes the sampling of early morning salivary cortisol following a midnight dexamethasone dose an easier method to diagnose Cushing disease in children. Transsphenoidal microadenomectomy is the treatment of choice for Cushing disease.

Exogenous glucocorticoids used to treat asthma or even overzealous use of topical corticosteroid ointments or creams may suppress growth. These iatrogenic causes of Cushing syndrome, if resolved early, may allow catch-up growth and so may not affect adult height. Thus, an accurate history of prior medications is important in diagnosis. Treatment of the underlying disorder (eg, transsphenoidal microadenomectomy for Cushing disease) will restore growth rate to normal (catch-up growth may occur initially) if epiphyseal fusion has not occurred, but adult height will depend on the length of the period of growth suppression.

5. PSEUDOHYPOPARATHYROIDISM

Pseudohypoparathyroidism type 1A (OMIM #103580) is a rare disorder consisting of a characteristic phenotype and biochemical signs of hypoparathyroidism (low serum calcium and high serum phosphate) in its classic form. Phosphate is elevated in states of deficient parathyroid hormone (PTH) action, compared to vitamin D deficiency where phosphate levels are decreased while PTH is elevated. Circulating PTH levels are elevated in pseudohypoparathyroidism, but target tissues fail to respond to exogenous PTH administration. Children with classic pseudohypoparathyroidism are short and overweight, with characteristic round facies and short fourth and fifth metacarpals. This constellation of physical findings is called Albright hereditary osteodystrophy, which may be expressed separately from the biochemical disorders (see later). Developmental delay is common. The condition is due to heterozygous loss of function mutations in the alpha subunit of the G_s protein transducer (*GNAS1* gene). When imprinted paternally, this results in a defect in the G-protein that couples PTH receptors to adenylyl cyclase due to inheritance of the defective maternal allele. Thus, patients with pseudohypoparathyroidism type 1A have a blunted rise of urinary cAMP in response to administration of PTH. Remarkably, a defect occurs in the same regulatory protein system affected in McCune-Albright syndrome, in which hyperactive endocrine events result (see Chapters 1 and 8). A rarer variant of this disorder (pseudohypoparathyroidism type 1B; OMIM #603233) appears to be due to mutations in noncoding regions of the *GNAS1* gene. Treatment with high-dose vitamin D or

1,25-dihydroxyvitamin D (calcitriol) and exogenous calcium as well as phosphate-binding agents (if needed) will help correct the biochemical defects and control hypocalcemic seizures in patients with pseudohypoparathyroidism.

Two remarkable cousins are reported with pseudohypoparathyroidism and premature Leydig cell maturation, both due to abnormalities in the same G protein. The defective protein was shown to be inactive at normal body temperature, leading to defective PTH activity at the level of the kidney and bone. However, it was hyperactive at the cooler temperatures in the scrotum, leading to ligand-independent activation of Leydig cell function.

Children with the phenotype of Albright hereditary osteodystrophy but with normal circulating levels of calcium, phosphate, and PTH have pseudopseudohypoparathyroidism (OMIM #612463). They require no calcium or vitamin D therapy (see Chapter 8).

6. DISORDERS OF VITAMIN D METABOLISM

Short stature and poor growth are features of rickets in its obvious or more subtle forms. The cause may be vitamin D deficiency due to inadequate oral intake, fat malabsorption, inadequate sunlight exposure, anticonvulsant therapy, and/or renal or hepatic disease. Classic findings of vitamin D-deficient rickets include bowing of the legs, chest deformities (rachitic rosary), and characteristic radiographic findings of the extremities associated with decreased serum calcium and phosphate levels and elevated serum alkaline phosphatase levels. There are two forms of hereditary vitamin D-dependent rickets. Autosomal recessive type 1 hypophosphatemic rickets (OMIM #241520) involves a renal 25 OH vitamin D 1-hydroxylase deficiency, and type 2 (OMIM #613312) involves an absent or defective vitamin D receptor. However, the most common type of rickets in the United States is X-linked hypophosphatemic rickets (OMIM #307800), a dominant genetic disorder affecting renal reabsorption of phosphate. It is associated with short stature, severe and progressive bowing of the legs (but no changes in the wrists or chest), normal or slightly elevated serum calcium, very low serum phosphate, and urinary phosphate wasting. Short stature is linked with rickets in other renal disorders associated with renal phosphate wasting. Examples include Fanconi syndrome (including cystinosis and other inborn errors of metabolism) and RTA.

When treatment is effective in these disorders (eg, vitamin D for vitamin D deficiency or alkali therapy for appropriate types of RTA), growth rates will improve. Replacement of vitamin D and phosphate is appropriate therapy for vitamin D-resistant rickets. It improves the bowing of the legs and leads to improved growth, although there is a risk of nephrocalcinosis with vitamin D treatment. This necessitates annual renal ultrasound examinations when patients are receiving vitamin D therapy.

In the Williams-Bureun syndrome of elfin facies, supravalvular aortic stenosis, and mental retardation with gregarious personality, patients have SGA and greatly reduced height in childhood and as adults; this disorder may include infantile hypercalcemia but is no longer considered a disorder of vitamin D metabolism because a genetic defect in the *elastin* gene at 7q11.23 occurs in most affected patients (#194050).

7. DIABETES MELLITUS

Growth in type 1 diabetes mellitus depends on the efficacy of therapy. Well-controlled diabetes mellitus is compatible with normal growth, whereas poorly controlled diabetes often causes slow growth. Liver and spleen enlargement in a poorly controlled, short, diabetic child is known as Mauriac syndrome, rarely seen now owing to improved diabetes care. Another factor that may decrease growth rate in children with type 1 diabetes mellitus is the increased incidence of Hashimoto thyroiditis in this population. Yearly thyroid function screening is advisable, especially as the peripubertal period approaches. GH concentrations are higher in children with diabetes, and this factor may play a role in the development of diabetic complications. IGF-I concentrations tend to be normal or low, depending on glucose control, but judging from the elevated GH noted earlier, the stimulation of IGF-I production by GH appears to be partially blocked in these children. Celiac disease may occur in 10% of children with type 1 diabetes and can itself lead to growth failure.

8. DIABETES INSIPIDUS

Polyuria and polydipsia due to inadequate vasopressin (central or neurogenic diabetes insipidus) or inability of the kidney to respond to vasopressin (nephrogenic diabetes insipidus) lead to poor caloric intake and decreased growth. With appropriate treatment (see Chapter 5), the growth rate should return to normal. Acquired neurogenic or central diabetes insipidus may indicate the development of a hypothalamic-pituitary tumor, and growth failure may be due to associated GH deficiency.

DIAGNOSIS OF SHORT STATURE (TABLES 6–3, 6–4, AND 6–5)

Evaluation of Short Stature

An initial decision must determine whether a child is pathologically short or simply distressed because height is not as close to the 50th percentile as desired by the patient or the parents. Performing unnecessary tests is expensive and may be a source of

TABLE 6–3 Basic diagnosis of short stature.

Medical History
Birth weight and gestational age
Prenatal substance abuse
Birth trauma or complications
Family heights
Age of puberty or menarche in family
Family history of or symptoms of chronic diseases
Dietary history
Physical Examination
Measured and charted height
Measured and charted weight and BMI
Signs of syndromes
Midline defects
Neurological examination
Findings of chronic disease

TABLE 6–4 The laboratory evaluation of short stature.

Test	Rationale
CBC	*Anemia*: nutritional, chronic disease, malignancy *Leukocytosis*: inflammation, infection *Leukopenia*: bone marrow failure syndromes *Thrombocytopenia*: malignancy, infection
ESR, CRP	Inflammation of infection, inflammatory diseases, malignancy
Chemistry panel (electrolytes, liver enzymes, BUN)	Signs of acute or chronic hepatic, renal, adrenal dysfunction; hydration and acid-base status
Carotene, folate, tissue transglutaminase antibody, or reflex celiac panel	Assess malabsorption; detect celiac disease
Urinalysis	Signs of renal dysfunction, hydration; renal tubular acidosis
Karyotype, candidate gene analysis, SNP analysis using microarrays on DNA chips, array-comparative genomic hybridization (array-CGH), or other genomic techniques	Evaluate for genetic syndromes
Cranial MRI imaging	Assesses hypothalamic-pituitary tumors (craniopharyngioma, glioma, germinoma) or congenital midline defects
Bone age	Determine physiologic maturation, and evaluate height potential
IGF-I, IGF BP3	Reflects growth hormone status or nutrition
Free thyroxine and TSH	Detects hypothyroidism
Prolactin	Elevated in hypothalamic dysfunction or destruction, suppressed in pituitary disease

Abbreviations: BUN, blood urea nitrogen; CBC, complete blood count; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; IGF, insulin like growth factor-binding protein 3; MRI, magnetic resonance imaging; TSH, thyroid-stimulating hormone.

TABLE 6–5 Growth hormone testing.

Growth hormone testing is last and only performed if no other diagnosis is found
Never get basal serum GH unless you suspect gigantism!
Stimulated tests:
L-Dopa
Clonidine
Arginine
Glucagon
Insulin-induced hypoglycemia (very dangerous)
GHRH
Growth hormone–releasing peptides (GRP)

long-term concern to the parents—a concern that could be avoided by appropriate reassurance. Alternatively, missing a diagnosis of pathologic poor growth may cause the patient to lose inches of final height or may allow progression of serious disease.

If a patient's stature, growth rate, or height adjusted for midparental height is sufficiently decreased to warrant evaluation, an orderly approach to diagnosis will eliminate unnecessary laboratory testing. The medical history will provide invaluable information as to the intrauterine course and toxin exposure and the possibility of birth trauma as well as an indication of other acute or chronic diseases and dietary intake. Birthweight and gestational age determine whether the child is SGA or AGA. The evaluation for dietary deficiencies and for symptoms of any chronic disease is important since almost any systemic disease or nutritional compromise can decrease growth rate (see Table 6–1). Review of past growth charts is important, but in this modern era, children often change doctors frequently, and these data may not be available. Asking whether the child has changed clothing sizes or shoe sizes is useful in the absence of any other data that may allow determination of growth rate. Heights of parents and age of puberty of parents are recorded although usually only a mother will recall her age of menarche while the father will not likely remember anything about his pubertal development (unless the father continued to grow after he left high school which might indicate constitutional delay in puberty). The height of siblings and specifically their percentile of height and whether they entered puberty at an appropriate time is important. The presence of chronic disease in the family is also noted in the history. Evaluation of psychosocial factors affecting the family and the relationship of parents and child can be carried out during the history-taking encounter. Often the diagnosis can be made at this point.

It goes without saying that accurate measurement of growth is essential. The physical examination requires determination of height as described earlier, and comparison is carried out with any previous data available. Measurement of weight and the calculation of BMI are performed so that neither obesity nor malnutrition is missed. If past heights are not available, a history of lack of change in clothing and shoe sizes or failure to lengthen skirts or pants may reflect poor growth. Questions about how the child's stature compares with that of his or her peers and whether the child's height has always had the same relationship to that of classmates is useful. One of the most important features of the evaluation process is to determine height velocity and compare the child's growth rate with the normal growth rate for age. Adjustment for midparental height is calculated and nutritional status determined. Arm span, head circumference, and US-LS ratio are measured. Physical examination is directed to uncover signs of chronic disease, CNS condition or a midline defect of the CNS that may be related to hypothalamic pituitary dysfunction. Physical stigmas of syndromes or systemic diseases are evaluated. Neurologic examination is essential.

Any clues to a diagnosis in the history or physical examination are pursued. However, if no historical or physical features lead to an etiology, laboratory examinations are performed (see Table 6–4). A list of chronic diseases causing short stature is presented in Table 6–1. Complete blood count and serum chemistry screening with electrolyte measurements may reveal anemia, abnormalities of hepatic or renal disease, glucose intolerance, acidosis, calcium disorder, or other electrolyte disturbances. Age-adjusted values must be used (eg, the normal ranges of serum alkaline phosphatase and phosphorus values are higher in children than in adults). An elevated sedimentation rate, low serum carotene, or a positive antinuclear, or tissue transglutaminase antibody determination (if IgA determination first eliminates the possibility of IgA deficiency providing a false negative test) may indicate connective tissue disease, Crohn's disease, celiac disease, or malabsorption syndrome. Serum TSH and free T_4 are important measurements to exclude existing thyroid disease. Urinalysis is done, with attention to specific gravity (to rule out diabetes insipidus) and ability to acidify urine (to evaluate possible RTA). Skeletal age evaluation does not alone provide a diagnosis; however, if the study shows delayed bone age, the possibility of constitutional delay in growth, hypothyroidism, or GH deficiency must be considered. The tests used for the diagnosis of GH deficiency are detailed earlier (see Table 6–5). If serum IGF-I is normal for age, classic GH deficiency or malnutrition is unlikely; if serum IGF-I is low, it must be considered in relation to skeletal age, nutritional status, and general health status before interpretation of the value can be made. Since IGF-I values are low under 2 or 3 years of age, the simultaneous measurement of IGFBP-3 is useful in infants. If either or both IGF-I and IGFBP-3 are low, the diagnosis may be GH deficiency if poor nutrition is ruled out. If GH deficiency or impairment is found or if there is another hypothalamic-pituitary defect, an MRI is indicated with particular attention to the hypothalamic-pituitary area to rule out a congenital defect or neoplasm in the area. The appearance of an ectopic location of the posterior pituitary on MRI is relatively frequent in congenital GH deficiency, as is a decreased pituitary volume or apparent interruption of the pituitary stalk. Serum gonadotropin and sex steroid determinations are performed in pediatric assays if puberty is delayed (see Chapter 15). Serum prolactin may be elevated in the presence of a hypothalamic disorder.

Celiac disease is quite common as a cause of gastrointestinal distress and/or short stature. Serum IgA levels and tissue transglutaminase antibody measurements are indicated in the evaluation of growth disorders. A karyotype is obtained in any short girl without another diagnosis to rule out Turner syndrome, especially if puberty is delayed or gonadotropins are elevated. If Turner syndrome is diagnosed, evaluation of thyroid function and determination of thyroid antibodies is also important.

Elevated urinary free cortisol (normal: <60 μg/m^2/24 h [<18.7 μmol/m^2/24 h]), elevated late night salivary cortisol, or abnormal dexamethasone suppression testing signifies Cushing syndrome.

If no diagnosis is apparent after all of the above have been considered and evaluated, more detailed procedures, such as provocative testing for GH deficiency, are indicated. It must be emphasized that a long and expensive evaluation is not appropriate until psychologic or nutritional factors are ruled out. Likewise, if a healthy-appearing child presents with borderline short stature, normal growth rate, and short familial stature, a period of observation may be more appropriate than laboratory tests.

TALL STATURE DUE TO NONENDOCRINE CAUSES

1. CONSTITUTIONAL TALL STATURE

A subject who has been taller than his or her peers through most of childhood is growing at a velocity within the normal range with a moderately advanced bone age, and has no signs of the disorders listed later, may be considered to be constitutionally advanced. Predicted final height will usually be in the normal adult range for the family.

Obesity in an otherwise healthy child will often, especially in the presence of a melanocortin 4 receptor (MC4R) mutation, lead to moderate advancement of bone age, slightly increased growth rate, and tall stature in childhood. Age of puberty will begin in the early range of normal, and adult stature will conform to genetic influences. Thus, an obese child without endocrine disease should be tall; short stature and obesity are worrisome.

2. FAMILIAL/GENETIC TALL STATURE

Children with exceptionally tall parents have a genetic tendency to reach a height above the normal range. The child will be tall for age and will grow at a normal to high normal rate. Bone age will be close to chronologic age, leading to a tall height prediction. Occasionally, children will be concerned about being too tall as adults. These worries are more common in girls and will often be of greater concern to the parents than to the patient. Adult height was limited in the past by promoting early epiphyseal closure with estrogen in girls or testosterone in boys, but such therapy is no longer considered appropriate. Testosterone therapy decreases HDL cholesterol levels. Acne fulminans may be caused by testosterone therapy and progression may occur, even after therapy has been withdrawn. Estrogen carries the theoretical risk of thrombosis, ovarian cysts, and galactorrhea, and recent report of decreased fertility. High-dose estrogen therapy is estimated to decrease predicted final height by as much as 4.5 to 7 cm but only if started 3 to 4 years before epiphyseal fusion. Such height-limiting therapy is extremely rare in the present era although, recently, long-acting somatostatin agonists have been used in an attempt to limit height in selected subjects. Counseling and reassurance are more appropriate.

3. SYNDROMES OF TALL STATURE

Cerebral Gigantism

The sporadic syndrome of rapid growth in infancy, prominent forehead, high-arched palate, sharp chin, and hypertelorism [Sotos syndrome (OMIM #117550)] is caused by mutation in the *nuclear receptor–binding SET domain protein 1* (*NSD1*) gene and is not associated with GH excess. Mentation is usually impaired. The growth rate decreases to normal in later childhood, but stature remains tall.

Marfan Syndrome

Marfan syndrome (OMIM #154700) is an autosomal dominant abnormality of connective tissue exhibiting variable penetrance.

The disorder is due to mutation of the *fibrillin 1* gene located on 15q21.1. This condition may be diagnosed by characteristic physical manifestations of tall stature, long thin fingers (arachnodactyly), hyperextension of joints, and superior lens subluxation. Pectus excavatum and scoliosis may be noted. Furthermore, aortic or mitral regurgitation or aortic root dilation may be present, and aortic dissection or rupture may ultimately occur. In patients with this syndrome, arm span exceeds height, and the US-LS ratio is quite low owing to long legs. Aortic root ultrasound and slit lamp ophthalmologic examinations are indicated.

Homocystinuria

Patients with homocystinuria (OMIM #236200) have an autosomal recessive deficiency of cystathionine beta-synthase (gene locus 21q22.3) and phenotypes similar to those of patients with Marfan syndrome. Additional features of homocystinuria include developmental delay, increased incidence of seizures, osteoporosis, inferior lens dislocation, and increased urinary excretion of homocystine with increased plasma homocystine and methionine but low plasma cystine. Thromboembolic phenomena may precipitate a fatal complication. This disease is treated by restricting dietary methionine and, in responsive patients, administering pyridoxine.

Beckwith-Wiedemann Syndrome

Patients with Beckwith-Wiedemann syndrome (OMIM #130650) demonstrate macrosomia (birth weight >90th percentile) in 88% of cases, increased postnatal growth, omphalocele in 80%, macroglossia in 97%, and hypoglycemia due to the hyperinsulinism of pancreatic hyperplasia in 63%. Other reported features include fetal adrenocortical cytomegaly and large kidneys with medullary dysplasia. The majority of patients occur in a sporadic pattern due to a mutation at 11p15.5, but analysis of some pedigrees suggests the possibility of familial patterns. There is a risk of Wilms tumor, hepatoblastoma, adrenal carcinoma, and gonadoblastoma in this condition.

XYY Syndrome

Patients with one (47,XYY) or more (48,XYYY) extra Y chromosomes achieve greater than average adult heights. They have normal birth lengths but higher than normal growth rates. Excess GH secretion has not been documented (see Chapter 14).

Klinefelter Syndrome

Patients with Klinefelter syndrome (see Chapters 12 and 14) tend toward tall stature, but this is not a constant feature.

TALL STATURE DUE TO ENDOCRINE DISORDERS

1. PITUITARY GIGANTISM

Pituitary gigantism is caused by excess GH secretion before the age of epiphyseal fusion. The increased GH secretion may be due

to somatotroph-secreting tumors or constitutive activation of GH secretion as is sometimes found in the McCune-Albright syndrome. Alternatively, it may result from excess secretion of GHRH. Patients—besides growing excessively rapidly—have coarse features, large hands and feet with thick fingers and toes, and often frontal bossing and large jaws.

Although this condition is quite rare, the findings appear similar to those observed in the more frequently diagnosed acromegaly (which occurs with GH excess after epiphyseal fusion). Thus, glucose intolerance or frank diabetes mellitus, hypogonadism, and thyromegaly are predicted. Treatment is accomplished by surgery (the transsphenoidal approach is used if the tumor is small enough), radiation therapy, or by medical therapy with a somatostatin analog.

2. SEXUAL PRECOCITY

Early onset of estrogen or androgen secretion leads to abnormally increased height velocity. Because bone age is advanced, there is the paradox of the tall child who, because of early epiphyseal closure, is short as an adult. The conditions include complete and incomplete sexual precocity (including virilizing congenital adrenal hyperplasia) (see Chapter 14).

3. THYROTOXICOSIS

Excessive thyroid hormone, due to endogenous overproduction or overtreatment with exogenous thyroxine, leads to increased growth, advanced bone age, and, if occurring in early life, craniosynostosis. If the condition remains untreated, adult height will be reduced due to early epiphyseal closure.

4. INFANTS OF DIABETIC MOTHERS

Birth weight and size in infants of diabetic mothers are quite usually high, although severely affected, poorly controlled mothers with type 1 diabetes may have infants with IUGR due to placental vascular insufficiency. Severe hypoglycemia and hypocalcemia are evident in the affected infants soon after birth. The appearance and size of such infants is so striking that women have been diagnosed with gestational diabetes as a result of giving birth to large affected infants. By 10 years of age, infants of diabetic mothers have an increased prevalence of obesity as well as insulin resistance and all of the comorbidities of the condition.

REFERENCES

Normal Growth

Greulich WW. Radiographic atlas of skeletal development off the hand and wrist. Stanford, CA: Stanford University Press; 1959. [repr. 1970]

Himes JH. Minimum time intervals for serial measurements of growth in recumbent length or stature of individual children. *Acta Paediatr*. 1999;88:120. [PMID: 10102141]

Kelly A, Winer KK, Kalkwarf H, et al. Age-based reference ranges for annual height velocity in US children. *J Clin Endocrinol Metab*. 2014;99:2104. [PMID: 24601728]

Lofqvist C, Andersson E, Gelander L, Rosberg S, Blum WF, Albertsson Wikland K. Reference values for IGF-I throughout childhood and adolescence: a model that accounts simultaneously for the effect of gender, age, and puberty. *J Clin Endocrinol Metab*. 2001;86:5870. [PMID: 11739455]

Munoz-Hoyos A, Molina-Carballo A, Augustin-Morales M, et al. Psychosocial dwarfism: psychopathological aspects and putative neuroendocrine markers. *Psych Res*. 2011;188:96. [PMID: 21071098]

Silventoinen K, Haukka J, Dunkel L, et al. Genetics of pubertal timing and its associations with relative weight in childhood and adult height: the Swedish Young Male Twins Study. *Pediatrics*. 2008;121:e885. [PMID: 18381517]

Styne DM. Growth. *Pediatric Endocrinology*. A Clinical Handbook Springer; 2016.

Towne B, Williams KD, Blangero J. Presentation, heritability, and genome-wide linkage analysis of the midchildhood growth spurt in healthy children from the Fels Longitudinal Study. *Hum Biol*. 2008;80:623. [PMID: 19728540]

http://www.cdc.gov/nchs/data/nhanes/growthcharts/set1/all.pdf.

Short Stature

Ahmad T, Garcia-Filion P, Borchert M, et al. Endocrinological and auxological abnormalities in young children with optic nerve hypoplasia: a prospective study. *J Pediatr*. 2006;148:78. [PMID: 16423602]

Albertsson-Wikland K, Aronson AS, Gustafsson J, et al. Dose-dependent effect of growth hormone on final height in children with short stature without growth hormone deficiency. *J Clin Endocrinol Metab*. 2008;93:4342. [PMID: 18728172]

Allen DB, Backeljauw P, Bidlingmaier M, et al. GH safety workshop position paper: a critical appraisal of recombinant human GH therapy in children and adults. *Eur J Endocrinol*. 2016;174:P1. [PMID: 26563978]

Backeljauw PF, Dattani MT, Cohen P, Rosenfeld RG. Disorders of growth hormone/insulin-like growth factor secretion and action. In: Sperling MA, ed. *Pediatric Endocrinology*. 4th ed. Philadelphia, PA: Elsevier; 2014:291-404.

Bondy CA, Turner Syndrome Study Group. Care of girls and women with Turner syndrome: a guideline of the Turner Syndrome Study Group. *J Clin Endocrinol Metab*. 2007;92:10. [PMID: 17047017]

Bullinger M, Kołtowska-Häggström M, Sandberg D, et al. Health-related quality of life of children and adolescents with growth hormone deficiency or idiopathic short stature—part 2: available results and future directions. *Horm Res*. 2009;72:74. [PMID: 19690424]

Chernausek SD, Backeljauw PF, Frane J, et al. Long-term treatment with recombinant insulin-like growth factor (IGF)-I in children with severe IGF-I deficiency due to growth hormone insensitivity. *J Clin Endocrinol Metab*. 2007;92:902. [PMID: 17192294]

Clayton PE, Cianfarani S, Czernichow P, et al. Management of the child born small for gestational age through to adulthood: a consensus statement of the International Societies of Pediatric Endocrinology and the Growth Hormone Research Society. *J Clin Endocrinol Metab*. 2007;92:804. [PMID: 17200164]

Cohen P, Rogol AD, Deal CL, et al. Consensus statement on the diagnosis and treatment of children with idiopathic short stature: a summary of the Growth Hormone Research Society, the Lawson Wilkins Pediatric Endocrine Society, and the European Society for Paediatric Endocrinology Workshop. *J Clin Endocrinol Metab*. 2008;93:4210. [PMID: 18782877]

Cole TJ, Wright CM. A chart to predict adult height from a child's current height. *Ann Hum Biol*. 2011;38:662. [PMID: 21767107]

Collett-Solberg PF. Update in growth hormone therapy of children. *J Clin Endocrinol Metab*. 2011;96:573. [PMID: 21378219]

Collett-Solberg PF, Misra M. The role of recombinant human insulin-like growth factor-I in treating children with short stature. *J Clin Endocrinol Metab*. 2008;93:10. [PMID: 18165284]

Cooke DW, Divall SA, Radovick S. Normal and aberrant growth. In: Melmed S, Polonsky KS, Larsen PR, Kronenberg HM, eds. *Williams Textbook of Endocrinology*. 12th ed. Philadelphia, PA: Saunders Elsevier; 2011:935-1053.

Davenport ML, Crowe BJ, Travers SH, et al. Growth hormone treatment of early growth failure in toddlers with Turner syndrome: a randomized, controlled, multicenter trial. *J Clin Endocrinol Metab*. 2007;92:3406. [PMID: 17595258]

David A, Hwa V, Metherell LA, et al. Evidence for a continuum of genetic, phenotypic, and biochemical abnormalities in children with growth hormone insensitivity. *Endocr Rev*. 2011;32:472. [PMID: 21525302]

de Graaff LC, Argente J, Veenma DC, et al. Genetic screening of a Dutch population with isolated GH deficiency (IGHD). *Clin Endocrinol (Oxf)*. 2009;70:742. [PMID: 18785993]

Deal CL, Tony M, Hoybye C, et al. Growth Hormone Research Society workshop summary: consensus guidelines for recombinant human growth hormone therapy in Prader-Willi syndrome. *J Clin Endocrinol Metab*. 2013;98:E1072. [PMID: 23543664]

Deodati A, Cianfarani S. Impact of growth hormone therapy on adult height of children with idiopathic short stature: systematic review. *BMJ*. 2011;342:c7157. [PMID: 21398350]

Divall SA, Radovick S. Growth hormone and treatment controversy; long term safety of rGH. *Curr Pediatr Rep*. 2013;1:128. [PMID: 23772352]

Donaldson MD, Gault EJ, Tan KW, Dunger DB. Optimising management in Turner syndrome: from infancy to adult transfer. *Arch Dis Child*. 2006;91:513. [PMID: 16714725]

GH Research Society. Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society. *J Clin Endocrinol Metab*. 2000;85:3990. [PMID: 11095419]

Gonc EN, Yordam N, Ozon A, et al. Endocrinological outcome of different treatment options in children with craniopharyngioma: a retrospective analysis of 66 cases. *Pediatr Neurosurg*. 2004;40:112. [PMID: 15367800]

Grote FK, van Dommelen P, Oostdijk W, et al. Developing evidence-based guidelines for referral for short stature. *Arch Dis Child*. 2008;93:212. [PMID: 17908714]

Hardin DS. GH improves growth and clinical status in children with cystic fibrosis—a review of published studies. *Eur J Endocrinol*. 2004;151(Suppl 1):S81. [PMID: 15339250]

Huguenin M, Trivin C, Zerah M, et al. Adult height after cranial irradiation for optic pathway tumors: relationship with neurofibromatosis. *J Pediatr*. 2003;142:699. [PMID: 12838200]

Keil MF, Stratakis CA. Pituitary tumors in childhood: update of diagnosis, treatment and molecular genetics. *Expert Rev Neurother*. 2008;8:563. [PMID: 18416659]

Kelly BP, Paterson WF, Donaldson MD. Final height outcome and value of height prediction in boys with constitutional delay in growth and adolescence treated with intramuscular testosterone 125 mg per month for 3 months. *Clin Endocrinol (Oxf)*. 2003;58:267. [PMID: 12608930]

Laron Z. The essential role of IGF-I: lessons from the long-term study and treatment of children and adults with Laron syndrome. *J Clin Endocrinol Metab*. 1999;84:4397. [PMID: 10599694]

Leschek EW, Rose SR, Yanovski JA, et al. Effect of growth hormone treatment on adult height in peripubertal children with idiopathic short stature: a randomized, double-blind, placebo-controlled trial. *J Clin Endocrinol Metab*. 2004;89:3140. [PMID: 15240584]

Martinelli CE Jr, Sader SL, Oliveira EB, Daneluzzi JC, Moreira AC. Salivary cortisol for screening of Cushing's syndrome in children. *Clin Endocrinol (Oxf)*. 1999;51:67. [PMID: 10468967]

Mauras N. Strategies for maximizing growth in puberty in children with short stature. *Pediatr Clin North Am*. 2011;58:1167. [PMID: 21981954]

Mullis PE. Genetics of growth hormone deficiency. *Endocrinol Metab Clin North Am*. 2007;36:17. [PMID: 17336732]

Oberfield SE, Sklar CA. Endocrine sequelae in survivors of childhood cancer. *Adolesc Med*. 2002;13:161. [PMID: 11841962]

Patel L, Dixon M, David TJ. Growth and growth charts in cystic fibrosis. *J R Soc Med*. 2003;96(Suppl 43):35. [PMID: 12906324]

Phillip M, Lebenthal Y, Lebl J, et al. European multicentre study in children born small for gestational age with persistent short stature: comparison of continuous and discontinuous growth hormone treatment regimens. *Horm Res*. 2009;71:52. [PMID: 19039237]

Poidvin A, Touze E, Ecosse E, et al. Growth hormone treatment for childhood short stature and risk of stroke in early adulthood. *Neurology*. 2014;83:780. [PMID: 25122206]

Ranke MB, Schweizer R, Lindberg A, et al. Insulin-like growth factors as diagnostic tools in growth hormone deficiency during childhood and adolescence: the KIGS experience. *Horm Res*. 2004;62(Suppl 1):17. [PMID: 15761228]

Ranke MB. Treatment of children and adolescents with idiopathic short stature. *Nat Rev Endocrinol*. 2013;9:325. [PMID: 23609338]

Rogol AD, Hayden GF. Etiologies and early diagnosis of short stature and growth failure in children and adolescents. *J Pediatr*. 2014;164(5 Suppl):S1. [PMID: 24731744]

Saenger P, Czernichow P, Hughes I, Reiter E. Small for gestational age: short stature and beyond. *Endocr Rev*. 2007;28:219. [PMID: 17322454]

Savage MO. Insulin-like growth factors, nutrition and growth. *World Rev Nutr Diet*. 2013;106:52. [PMID: 23428681]

Simon D, Leger J, Carel JC. Optimal use of growth hormone therapy for maximizing adult height in children born small for gestational age. *Best Prac Res Clin Endocrinol Metab*. 2008;22:525. [PMID: 18538291]

Stanley TL, Levitsky LL, Grinspoon SK, Misra M. Effect of body mass index on peak growth hormone response to provocative testing in children with short stature. *J Clin Endocrinol Metab*. 2009;94:4875. [PMID: 19890023]

Styne DM, Grumbach MM, Kaplan SL, Wilson CB, Conte FA. Treatment of Cushing's disease in childhood and adolescence by transsphenoidal microadenomectomy. *N Engl J Med*. 1984;310:889. [PMID: 6321987]

Tanner JM, Davies PS. Clinical longitudinal standards for height and height velocity for North American children. *J Pediatr*. 1985;107:317. [PMID: 3875704]

van Dijk M, Bannink EM, van Pareren YK, et al. Risk factors for diabetes mellitus type 2 and metabolic syndrome are comparable for previously growth hormone-treated young adults born small for gestational age (SGA) and untreated short SGA controls. *J Clin Endocrinol Metab*. 2007;92:160. [PMID: 17062774]

Wickman S, Kajantie E, Dunkel L. Effects of suppression of estrogen action by the p450 aromatase inhibitor letrozole on bone mineral density and bone turnover in pubertal boys. *J Clin Endocrinol Metab*. 2003;88:3785. [PMID: 11888840]

Wit JM. Definition and subcategorization of idiopathic short stature: between consensus and controversy. *Horm Res Paediatr*. 2011;76(Suppl 3):3. [PMID: 21912144]

Wit JM, Hero M, Nunez SB. Aromatase inhibitors in pediatrics. *Nat Rev Endocrinol*. 2012;8:135. [PMID: 22024975]

Wit JM, Oostdijk W, Losekoot M. Spectrum of insulin-like growth factor deficiency. *Endocr Dev*. 2012;23:30. [PMID: 23182818]

Wolfgram PM, Carrel AL, Allen DB. Long-term effects of recombinant human growth hormone therapy in children with Prader-Willi syndrome. *Curr Opin Pediatr*. 2013;25(4):509. [PMID: 23782572]

Yanovski JA, Rose SR, Municchi G, et al. Treatment with a luteinizing hormone-releasing hormone agonist in adolescents with short stature. *N Engl J Med*. 2003;348:908. [PMID: 12621135]

Tall Stature

Eugster EA, Peskovitz OH. Gigantism. *J Clin Endocrinol Metab*. 1999;84:4379. [PMID: 10599691]

Leventopoulos G, Kitsiou-Tzeli S, Kritikos K, et al. A clinical study of Sotos syndrome patients with review of the literature. *Pediatr Neurol*. 2009;40:357. [PMID: 19380072]

Thomsett MJ. Referrals for tall stature in children: a 25-year personal experience. *J Paediatr Child Health*. 2009;45:58. [PMID: 19208068]

This page intentionally left blank

The Thyroid Gland

David S. Cooper, MD and
Paul W. Ladenson, MD (Oxon)., MD

Ab	antibody		**rhTSH**	recombinant human TSH
ClO$_4^-$	perchlorate ion		**rT$_3$**	reverse triiodothyronine
DIT	diiodotyrosine		**RTH**	resistance to thyroid hormones
FNAB	fine-needle aspiration biopsy		**RXR**	retinoid X receptor
FSH	follicle-stimulating hormone		**SMRT**	silencing mediator for retinoic and thyroid hormone receptors
GRTH	generalized resistance to thyroid hormone		**T$_4$**	thyroxine
hCG	human chorionic gonadotropin		**T$_3$**	triiodothyronine
LH	luteinizing hormone		**TBG**	thyroxine-binding globulin
MCT	monocarboxylate transporter		**TBPA**	transthyretin or thyroxine-binding prealbumin
MIT	monoiodotyrosine		**TG**	thyroglobulin
NADPH	nicotinamide adenine dinucleotide phosphate		**TPO**	thyroid peroxidase
NCoR	nuclear receptor corepressor		**TR**	thyroid hormone receptor
NIS	sodium-iodide symporter		**TRH**	thyrotropin-releasing hormone
OATP1C1	organic anion transporting polypeptide		**TSH**	thyroid-stimulating hormone (thyrotropin)
PTU	propylthiouracil		**WHO**	World Health Organization
RAIU	radioactive iodine uptake			

The thyroid gland is the body's largest single organ specialized for endocrine hormone production. Its function is to secrete an appropriate amount of the thyroid hormones, primarily 3,5,3′,5′-l-tetraiodothyronine (thyroxine, T$_4$), and a lesser quantity of 3,5,3′-l-triiodothyronine (T$_3$), which arises mainly from the subsequent extrathyroidal deiodination of T$_4$. In target tissues, T$_3$ interacts with nuclear T$_3$ receptors that are, in turn, bound to special nucleotide sequences in the promoter regions of genes that are positively or negatively regulated by thyroid hormone. The thyroid hormones promote normal fetal and childhood growth and central nervous system development; regulate heart rate and myocardial contraction and relaxation; affect gastrointestinal motility and renal water clearance; and modulate the body's energy expenditure, heat generation, weight, and lipid metabolism. In addition, the thyroid contains parafollicular or C cells that produce calcitonin, a 32-amino-acid polypeptide that inhibits bone resorption, but has no apparent physiologic role in humans.

However, calcitonin is clinically important as a tumor marker produced by medullary thyroid cancers that arise from these cells (see Chapter 8).

EMBRYOLOGY, ANATOMY, AND HISTOLOGY

The thyroid gland originates in the embryo as a mesodermal invagination in the pharyngeal floor at the foramen cecum, from which it descends anterior to the trachea and bifurcates, forming two lateral lobes, each measuring approximately 4 cm in length, 2 cm in width, and 1 cm in thickness in adulthood. Ectopic thyroid tissue can be present anywhere along or beyond this thyroglossal duct, from the tongue base (lingual thyroid) to the mediastinum. The thyroglossal duct may also give rise to midline
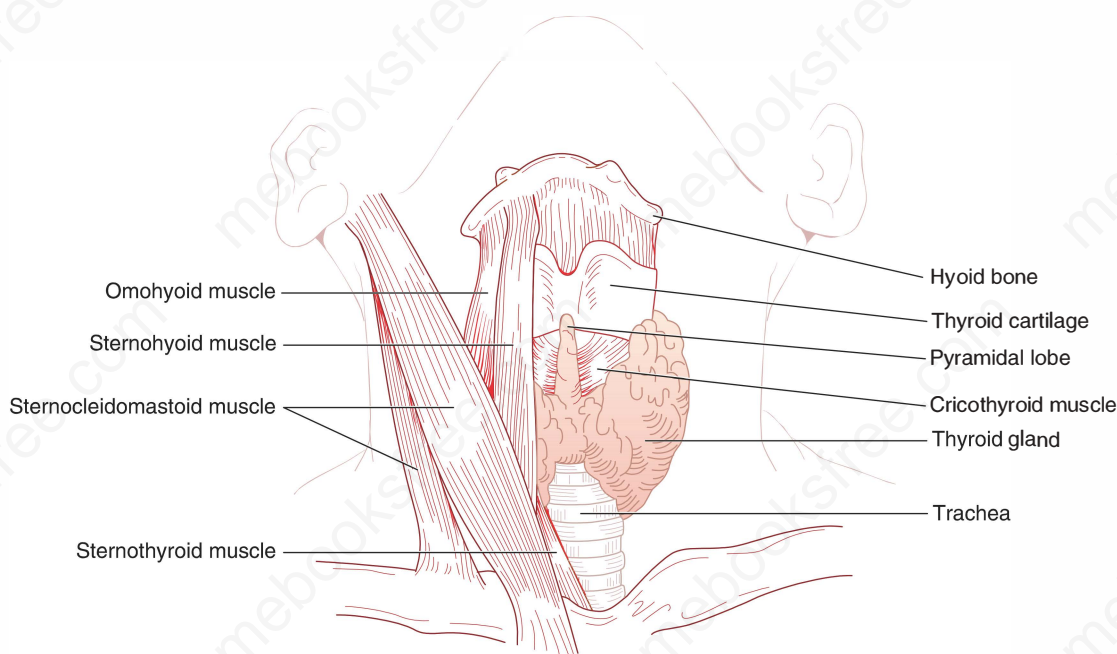
FIGURE 7–1 Gross anatomy of the human thyroid gland (anterior view).

cysts lined with squamous epithelium, which can remain asymptomatic, or become infected or give rise to thyroid tumors. The caudal end of the thyroglossal duct forms the pyramidal lobe of the thyroid, which can become palpable in conditions causing diffuse thyroid inflammation or stimulation (Figure 7–1).

Because the posterior thyroid capsule is bound to the pretracheal fascia, the gland normally rises and falls with deglutition, facilitating its inspection and palpation. When the thyroid enlarges, posterior and downward growth of goiter can extend into the superior mediastinum (substernal goiter), whereas upward extension of the thyroid gland is limited by the attachment of the sternothyroid muscle to the thyroid cartilage. The thyroid gland also has clinically important anatomical relationships to the recurrent laryngeal nerves, which course behind the gland, and two pairs of parathyroid glands that usually lie behind the upper and middle portions of the thyroid lobes. The thyroid is also draped around the trachea and the posterior margins of its lobes abut the esophagus. All of these structures can be compressed by gland enlargement, invaded by thyroid malignancies, or injured in the course of thyroid surgery (Figure 7–2).

The thyroid gland has a rich blood supply (Figure 7–3), which can be increased in hyperthyroidism,[1] giving rise to an audible whooshing sound (bruit) or even a palpable vibration (thrill). Microscopically, thyrocytes form hollow spheres (follicles) that surround a central lumen containing an aggregation of iodinated thyroglobulin (TG) (colloid) that represents the gland's hormone stores (Figure 7–4).

---

[1]In this chapter, the term "hyperthyroidism," which strictly speaking refers only to thyroid hormone excess due to overproduction of thyroid hormone by the gland, is sometimes used interchangeably with "thyrotoxicosis," which indicates all states of thyroid hormone excess.

PHYSIOLOGY

STRUCTURE AND SYNTHESIS OF THYROID HORMONES

The thyroid hormones are iodinated thyronines, which consist of two tyrosine moieties joined by an ether linkage (Figure 7–5). The follicular cells of the thyroid gland are specialized in their ability to synthesize the large hormonal precursor protein TG, concentrate iodide intracellularly from the circulation, and express a receptor that binds thyroid-stimulating hormone (thyrotropin, TSH), which promotes thyrocytes' growth and biosynthetic functions.

IODINE METABOLISM

Iodine[2] is a key structural component of thyroid hormones. Consequently, it is an essential micronutrient consumed in food or water as iodide or iodate, which is converted to iodide in the stomach. The World Health Organization (WHO) recommends a daily dietary iodine intake of 150 μg for adults, 200 μg for pregnant and lactating women, and 90 μg for children. Because most iodide is excreted by the kidneys, urinary iodide excretion is an excellent index of dietary intake.

Over millennia, iodine has been leached from the soil in many mountainous and inland regions of the world. Consequently, according to WHO, dietary iodine deficiency, defined as a daily

---

[2]In this chapter, the words "iodine," referring to the uncharged element, and "iodide," referring to its negatively charged anion form, are used interchangeably.
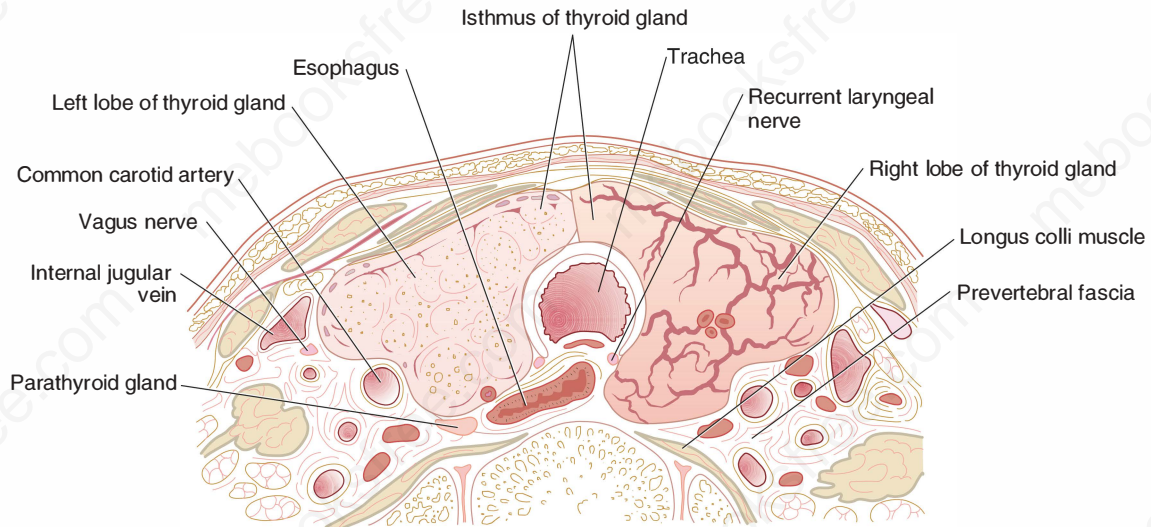
FIGURE 7–2 Cross-section of the neck at the level of T$_1$, showing thyroid relationships. (Reproduced with permission from Lindner HH. *Clinical Anatomy*. New York: McGraw-Hill Education; 1989.)

iodine intake less than 100 μg/d, affects an estimated two billion people, which is about one-third of the world's population. When iodide intake is less than 50 μg/d, a normal-sized thyroid cannot sustain adequate hormone production, with resulting gland enlargement (goiter) and, ultimately, hypothyroidism. The consequences of endemic dietary iodine deficiency are especially devastating for the developing fetus and children, who require thyroid hormone for normal neurologic development and growth. In the North American diet, iodine is principally derived from iodized

salt, iodate preservatives in baked goods, dairy products containing traces of iodophore antibacterial agents used in milk collection, food coloring, and seafood.

Thyrocytes abundantly express the sodium-iodide symporter (Na-I$^-$ symporter; NIS), which spans the cells' basal membranes and actively transports iodide from the blood. The thyroid gland concentrates and uses only a fraction of the iodide supplied to it for hormone synthesis, and the remainder returns to the extracellular fluid pool. Consequently, the normal fractional uptake of
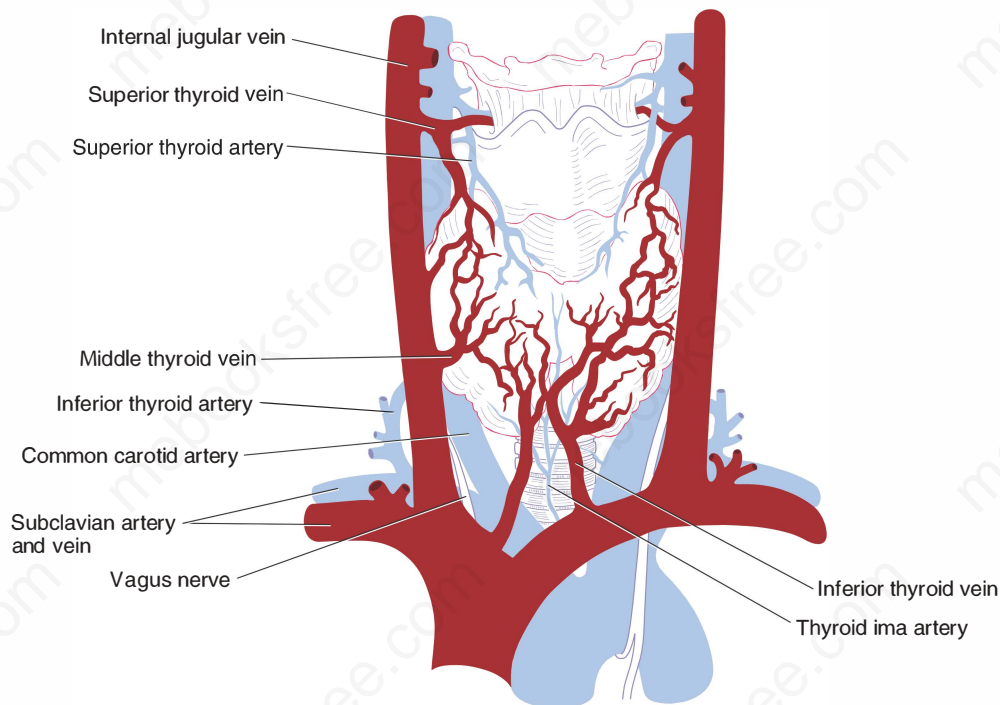
FIGURE 7–3 Arteries and veins related to the thyroid gland. (Reproduced with permission from Lindner HH. *Clinical Anatomy*. New York: McGraw-Hill Education; 1989.)
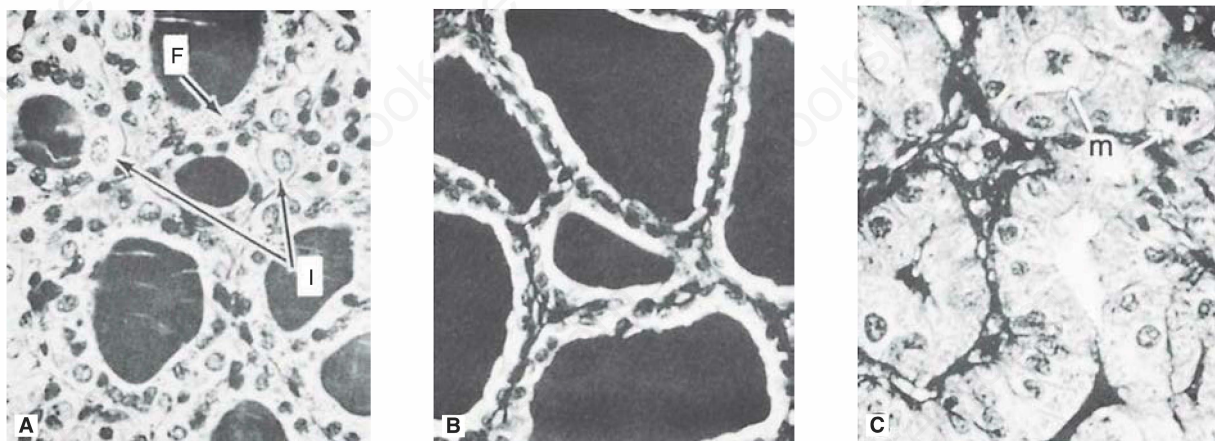
FIGURE 7–4   **A.** Normal rat thyroid. A single layer of cuboidal epithelial cells surrounds PAS-positive material in the follicular space (colloid). The larger, lighter-staining cells indicated by the arrows (I) are C cells that produce calcitonin. The follicular cells form an epithelial layer that surrounds the colloid (F). **B.** Inactive rat thyroid several weeks after hypophysectomy. The follicular lumens are larger and the follicular cells flatter. **C.** Rat thyroid under intensive TSH stimulation. The animal was fed an iodine-deficient diet and injected with propylthiouracil for several weeks. Little colloid is visible. The follicular cells are tall and columnar. Several mitoses (m) are visible. (Reproduced with permission from Greep RO, Weiss L. *Histology*. New York: McGraw-Hill Education;1973.)
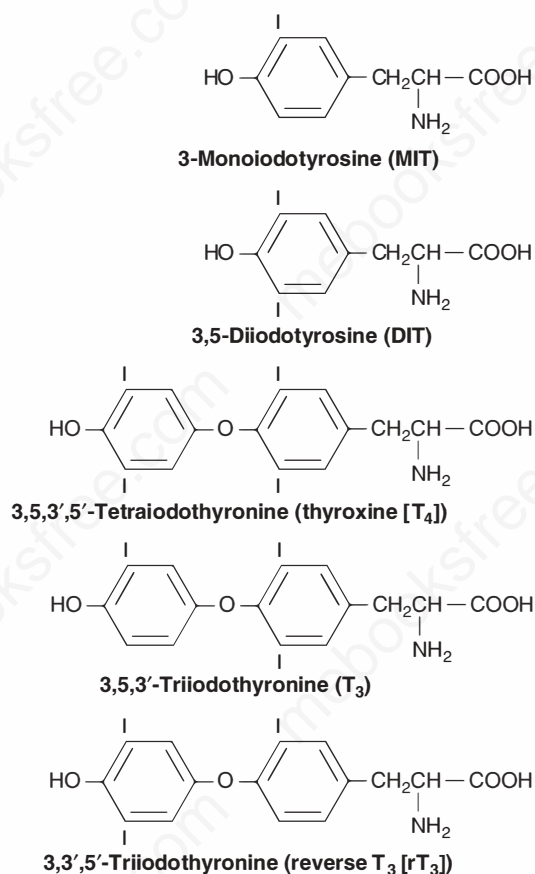
FIGURE 7–5   Structure of thyroid hormones and related compounds. (Reproduced with permission from Murray RK, Granner DK, Mayes PA, et al. *Harper's Biochemistry*, 24th ed. New York: McGraw-Hil Education; 1996.)

iodide, which can be quantified with a radioactive iodine tracer, is approximately 10% to 30% after 24 hours. Because of this active concentrating mechanism and the subsequent organification of intracellular iodide, the intrathyroidal pool of iodine, 8 to 10 mg in the form of stored thyroid hormones and iodinated tyrosines, provides a buffer in the event of temporary dietary iodine deficiency.

THYROID HORMONE SYNTHESIS AND SECRETION

Synthesis of T_4 and T_3 by the thyroid gland involves six major steps: (1) active transport of iodide across the basement membrane into the thyroid cell (trapping); (2) oxidation of iodide and iodination of tyrosyl residues in TG (organification); (3) linking pairs of iodotyrosine molecules within TG to form the iodothyronines T_3 and T_4 (coupling); (4) pinocytosis and then proteolysis of TG with release of free iodothyronines and iodotyrosines into the circulation; (5) deiodination of iodotyrosines within the thyroid cell, with conservation and reuse of the liberated iodide; and (6) intrathyroidal 5′-deiodination of T_4 to T_3.

Thyroid hormone synthesis requires that NIS, TG, and the enzyme thyroid peroxidase (TPO) all be present, functional, and uninhibited. This process is summarized in Figures 7–6 and 7–7.

Thyroglobulin

TG is a large glycoprotein molecule (MW 660,000 kD), composed of two subunits, each containing 5496 amino acids. TG includes approximately 140 tyrosyl residues, but only four tyrosyl sites in each molecule are sterically oriented for effective hormonogenesis. The iodine content of TG can vary from 0.1% to 1% by weight. In TG containing 0.5% iodine, for example, there are approximately three molecules of T_4 and one molecule of T_3.
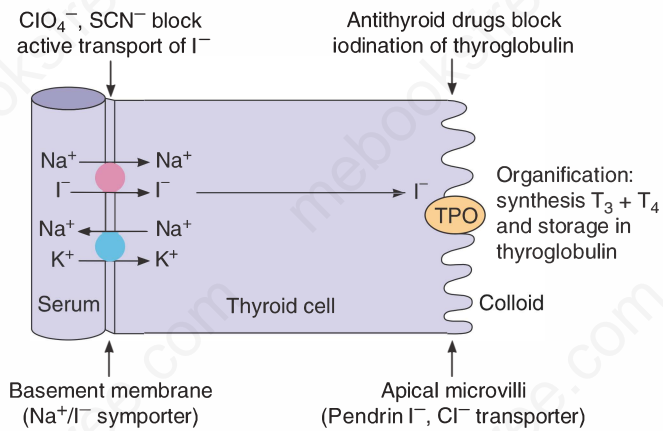
FIGURE 7–6 The iodide transporter in the thyroid cell. The pink solid circle represents the Na^+/I^- symporter actively transporting I^- into the cell; the blue solid circle represents Na^+-K^+ ATPase supplying the ion gradient which drives the reaction. I^- is transported across the apical membrane by pendrin. Hormone synthesis takes place in the colloid at the colloid–apical membrane, catalyzed by thyroperoxidase (TPO).

The *TG* gene, which resides on the long arm of chromosome 8, contains approximately 8500 nucleotides, which encode the prethyroglobulin protein monomer, including a 19-amino acid signal peptide. TSH regulates expression of the *TG* gene. After TG mRNA is translated in the rough endoplasmic reticulum (RER), the protein is glycosylated during transport through the Golgi apparatus (see Figure 7–7), where TG dimers are incorporated into exocytic vesicles. These vesicles then fuse with the cell's apical basement membrane, from which they are released into the follicular lumen. There, at the apical–colloid border, tyrosine residues in TG are iodinated.

Iodide Transport

Iodide (I^-) is transported across thyrocytes' basal membrane by the NIS. Membrane-bound NIS, which derives its energy from a Na^+-K^+ ATPase, allows the human thyroid gland to maintain a concentration of free iodide 30 to 40 times higher than that in plasma. NIS action is stimulated physiologically by TSH and pathophysiologically by the TSH receptor–stimulating antibody (Ab) of Graves disease. Salivary, gastric, and breast tissues also express NIS and concentrate iodide to a lesser extent than the thyroid, but these other tissues do not organify or store iodide, and their NIS activities are not stimulated by TSH. However, these tissues may concentrate enough iodine to be visualized on radioiodine scans and irradiated with high-dose 131-I therapy, causing radiation sialadenitis and gastritis.

Large amounts of iodide suppress both NIS activity and *NIS* gene expression, representing mechanisms of iodine autoregulation (see later). The perchlorate ion (ClO_4^-) competes with iodide for NIS; perchlorate has been used to treat hyperthyroidism and
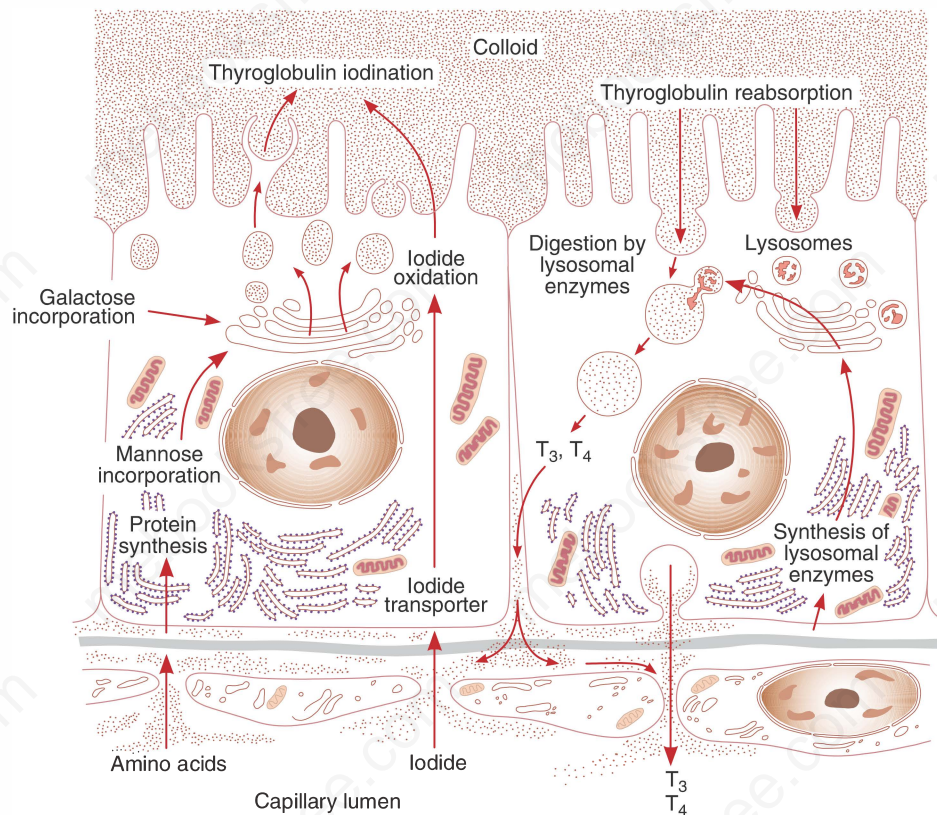
FIGURE 7–7 Processes of synthesis and iodination of thyroglobulin (left) and its reabsorption and digestion (right). These events occur in the same cell. (Reproduced with permission from Junqueira LC, Carneiro J, Kelley R. *Basic Histology*, 7th ed. New York: McGraw-Hill Education; 1992.)

has the potential to be an environmental inhibitor of thyroid function. NIS can also concentrate pertechnetate (TcO_4^-) into thyroid cells, facilitating use of the radionuclide sodium pertechnetate ($^{99m}TcO_4^-$) for visualization of the thyroid gland and quantification of its trapping activity.

At the thyrocyte's apical border, a second iodide transport protein, pendrin, transports iodide to the membrane–colloid interface, where it becomes a substrate for thyroid hormonogenesis (see Figures 7–6 and 7–7). Mutations in the *pendrin* (PDS or SLC26A4) gene impairing its function cause a syndrome of goiter and hearing loss acquired in infancy or early childhood (Pendred syndrome).

Thyroid Peroxidase

TPO, a membrane-bound glycoprotein (MW 102 kD) containing a heme moiety, catalyzes both iodide oxidation and covalent linkage of iodine to the tyrosine residues of TG. *TPO* gene expression is stimulated by TSH. After TPO is synthesized in the RER, it is inserted into the RER's cisternal membranes, from which it is transferred to the apical cell surface through Golgi elements and exocytic vesicles. Here, at the cell–colloid interface, TPO facilitates both iodination and coupling of the tyrosine residues within TG.

Iodination of Thyroglobulin

Within the thyroid cell, at the apical–colloid interface, iodide is rapidly oxidized by locally produced hydrogen peroxide in a reaction catalyzed by TPO; the resulting active iodide intermediate is bound to tyrosyl residues in TG. The required hydrogen peroxide is probably generated by a nicotinamide adenine dinucleotide phosphate (NADPH) oxidase in the presence of calcium cations, a process that is also stimulated by TSH.

Coupling of Iodotyrosyl Residues in Thyroglobulin

The coupling of iodotyrosyl residues in TG is also catalyzed by TPO. This is believed to be an intramolecular process involving oxidation of two iodotyrosyl residues brought into proximity by the tertiary and quaternary structures of TG, their linkage as a quinol ether intermediate, and splitting of the quinol ether to form an iodothyronine (Figure 7–8). Within the TG molecule, two molecules of diiodotyrosine (DIT) couple to form T_4, and monoiodotyrosine (MIT) and DIT molecules couple to form T_3.

The thiocarbamide drugs, including methimazole, carbimazole, and propylthiouracil (PTU), are competitive inhibitors of TPO. Their resulting ability to block thyroid hormone synthesis (Figure 7–9) makes them useful in treatment of hyperthyroidism.

Proteolysis of Thyroglobulin and Thyroid Hormone Secretion

The processes of TG proteolysis and secretion in thyroid hormonogenesis are illustrated in Figure 7–7. At the thyrocyte's apical
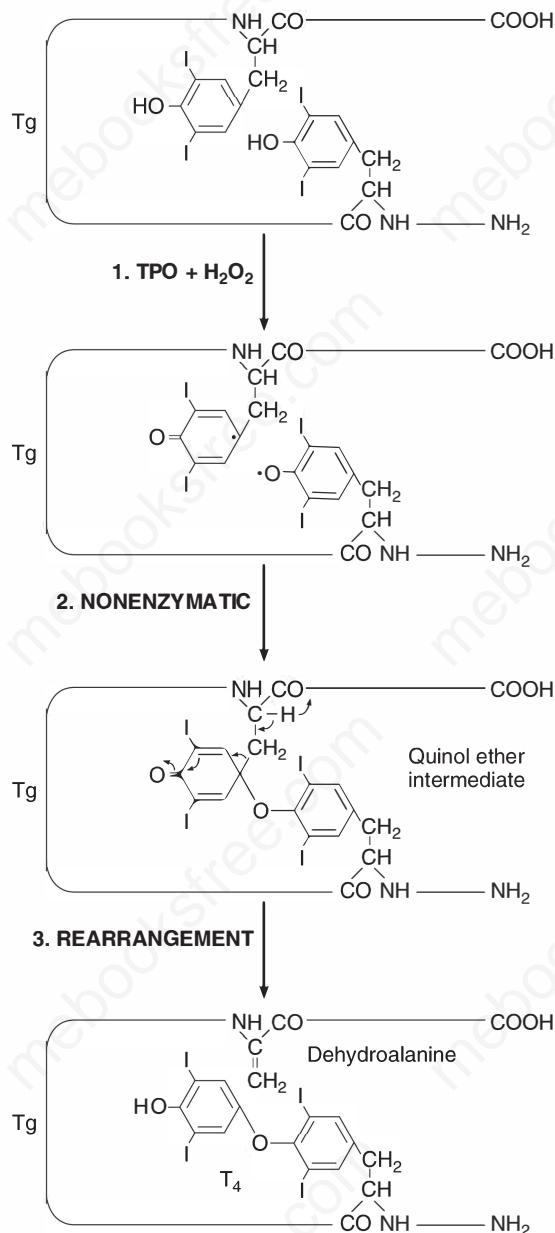
FIGURE 7–8 Hypothetical coupling scheme for intramolecular formation of T_4 within the thyroglobulin molecule. The major hormonogenic site at tyrosyl residue 5 is indicated. (Reproduced with permission from Braverman LE, Utiger RD. *Werner and Ingbar's The Thyroid*, 7th ed. Philadelphia: Lippincott Williams & Wilkins; 1996.)

membrane, colloid is engulfed in vesicles by pinocytosis and absorbed into the cell. Lysosomes containing proteolytic enzymes then fuse with the colloid vesicle. This releases T_4 and T_3, as well as inactive iodotyrosines, peptides, and individual amino acids. The biologically active thyroid hormones T_4 and T_3 enter the circulation; DIT and MIT are deiodinated and their iodide conserved. Thyroid hormone secretion is stimulated by TSH and inhibited by excess iodide (see later) and lithium. Intact TG is also normally released from thyroid cell and circulates in blood.
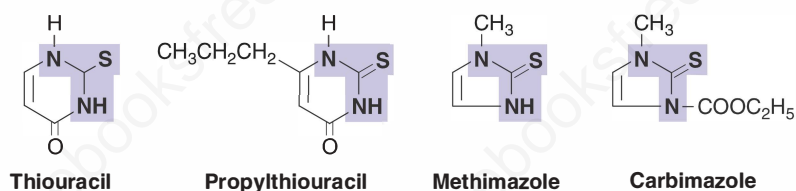
FIGURE 7–9 Thiocarbamide inhibitors of thyroidal iodide organification.

The serum TG concentration is markedly increased in a number of thyroid conditions, including thyroiditis, nodular goiter, and Graves disease. Because TG is also synthesized by most malignancies arising from thyroid epithelium, such as papillary and follicular thyroid cancers, it is a useful circulating tumor marker.

Intrathyroidal Deiodination

MIT and DIT formed during the synthesis of thyroid hormone are deiodinated by intrathyroidal deiodinase, an NADPH-dependent flavoprotein found in mitochondria and microsomes that acts on the iodotyrosines MIT and DIT, but not on T_3 and T_4. Most of the iodide released is reused for hormone synthesis, and only a small amount normally leaks out of the thyroid gland (Figure 7–10).

The 5′-deiodinase that converts T_4 to T_3 in peripheral tissues is also found in the thyroid gland. When there is iodide deficiency and in various hyperthyroid states, the activity of this enzyme increases the amount of T_3 secreted by the gland, increasing the metabolic efficiency of hormone synthesis.

ABNORMALITIES IN THYROID HORMONE SYNTHESIS AND RELEASE

Dietary Iodine Deficiency and Inherited Defects

A very low-iodine diet and inherited defects in genes encoding the proteins required for thyroid hormone biosynthesis (dyshormonogenesis) can both result in insufficient hormone production. In response to limited intrathyroidal iodine content or hormone production, the gland increases the ratios of MIT-to-DIT within TG as well as the proportion of secreted T_3 relative to T_4. The hypothalamic-pituitary-thyroid axis also responds to thyroid hormone deficiency by increasing TSH secretion. Consequently, affected individuals typically present with thyroid gland enlargement (goiter), which may be sufficient to compensate for inefficient thyroid hormone production; but if not, they develop hypothyroidism. Severely affected neonates and infants can suffer the irreversible effects of thyroid hormone deficiency on
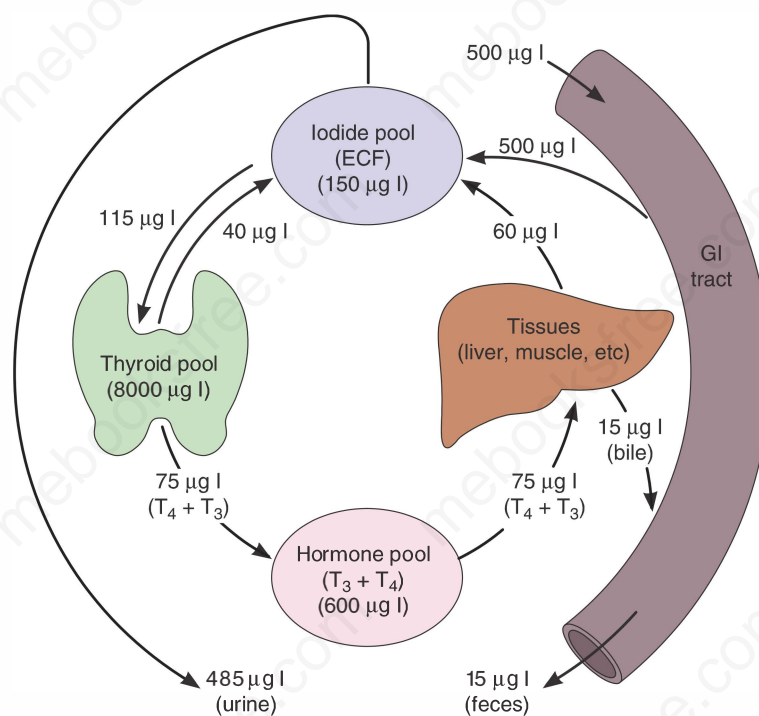
FIGURE 7–10 Iodine metabolism. The values indicated are representative of those that might be found in a healthy subject ingesting 500 μg of iodine a day. The actual iodine intake varies considerably among different individuals.

development that result in cretinism. Specific inherited disorders are described in more detail in the section Nontoxic Goiter, later.

Effects of Iodine Excess on Hormone Biosynthesis

Although iodide is an essential substrate for thyroid hormone production, excess iodide actually inhibits three steps in thyroid hormone production: iodide trapping, TG iodination (the Wolff-Chaikoff effect), and thyroid hormone release from the gland. These inhibitory actions are transient, and the normal thyroid gland *escapes* after 10 to 14 days from these effects of excess iodide. These autoregulatory effects of iodide insulate physiologic thyroid function from short-term fluctuations in iodine intake.

These actions of excess iodide also have important clinical implications, sometimes causing iodine-induced thyroid dysfunction. If the thyroid is affected by autoimmune thyroiditis or certain inherited forms of dyshormonogenesis, it may be incapable of escaping from sustained iodide-induced inhibition of gland function, and hypothyroidism can develop. Conversely, an iodide load can induce hyperthyroidism (Jod-Basedow effect[3]) in some patients with multinodular goiter, latent Graves disease, and rarely in individuals with thyroid glands that appear otherwise normal. The inhibitory effects of pharmacological iodide doses make iodide a useful treatment for short-term control of hyperthyroidism, such as before surgery. Iodide can also prevent thyroid gland exposure to radiation injury when populations are exposed to environmental radioiodine, such as after nuclear power plant accidents.

THYROID HORMONE TRANSPORT

Both thyroid hormones circulate in blood bound to plasma proteins; only 0.04% of T_4 and 0.4% of T_3 are unbound or free, and consequently, available for entry and action in target tissues (Figure 7–11). There are three major thyroid hormone transport proteins: thyroxine-binding globulin (TBG); transthyretin, formerly called thyroxine-binding prealbumin (TBPA), and albumin (Figure 7–12). The plasma protein binding permits blood delivery of the iodothyronines, which are otherwise poorly soluble in water. It also creates a large circulating thyroid hormone pool with a stable 7-day plasma half-life and ensures the homogeneous distribution of thyroid hormones in target tissues.

Thyroxine-Binding Globulin

TBG is a liver-derived glycoprotein member of the SERPIN family of serine antiproteases composed of a single 54-kD polypeptide chain, to which are attached four carbohydrate chains normally containing approximately 10 sialic acid residues. Each TBG molecule has a single binding site for T_4 or T_3. The serum concentration of TBG is 15 to 30 μg/mL (280-560 nmol/L), and its high-binding affinity for T_4 and T_3 allows it to carry about 70% of circulating thyroid hormones.

[3]Jod is German for iodine; Carl Adolph von Basedow was one of the first physicians to describe hyperthyroidism.
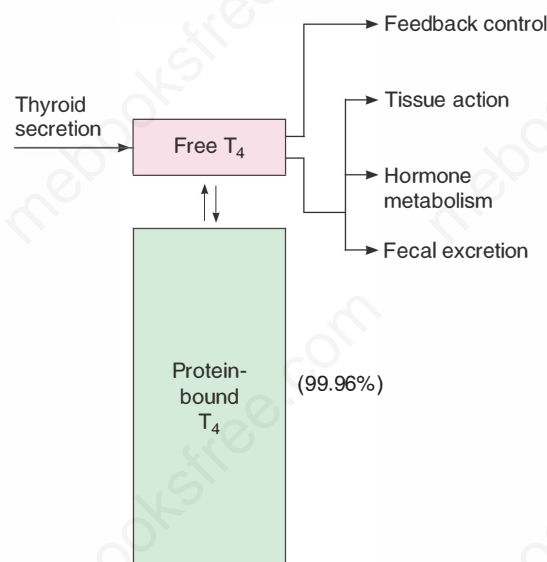
FIGURE 7–11 Representation of free T_4 (and free T_3) as the biologically active hormones at the level of the pituitary and the peripheral tissues. Most of the thyroid hormones circulating in plasma are protein-bound and have no biologic activity. This pool of bound hormone is in equilibrium with the free hormone pool. (Reproduced with permission from DeGroot LJ, Stanbury JB. *The Thyroid and Its Diseases*, 4th ed. Philadelphia: Elsevier; 1975.)

TBG and its binding of the thyroid hormones can be altered by congenital TBG derangements, certain physiologic and pathophysiologic circumstances, and several drugs. TBG deficiency occurs with a frequency of 1:5000 live births, with a number of variants described in several ethnic and racial groups. It is an X-linked recessive trait that, consequently, is much more commonly expressed in males. Despite low circulating total T_4 and T_3 levels in affected individuals, free-hormone levels are normal, and these patients remain euthyroid. Congenital TBG deficiency is often associated with congenital corticosteroid-binding globulin deficiency (see Chapter 9). Conversely, congenital TBG excess, which is rare, is characterized by elevated total T_4 and T_3 concentrations in blood, but normal free-hormone levels and a euthyroid clinical state. Pregnancy, estrogen-secreting tumors, and estrogen therapy all increase the sialic acid content of the TBG molecule, resulting in decreased metabolic clearance and elevated serum TBG levels. TBG levels can decrease with accelerated clearance in nephrotic syndrome and protein-losing enteropathy. TBG can decrease in major systemic illness due to cleavage by leukocyte proteases and reduction in TBG's binding affinity for the thyroid hormones. Both of these effects lower the serum total thyroid hormone concentrations in sick patients (Table 7–1). TBG levels can also decrease chronic thyrotoxicosis, hypercortisolism, and acromegaly. Certain drugs can decrease (androgenic steroids, glucocorticoids, danazol, L-asparaginase) or increase (estrogens, 5-fluourouracil) the plasma TBG concentration. Other drugs (eg, salicylates, high-dose phenytoin, and intravenous furosemide) can bind to TBG, displacing T_4 and T_3. In this circumstance, the hypothalamic-pituitary-thyroid axis (discussed
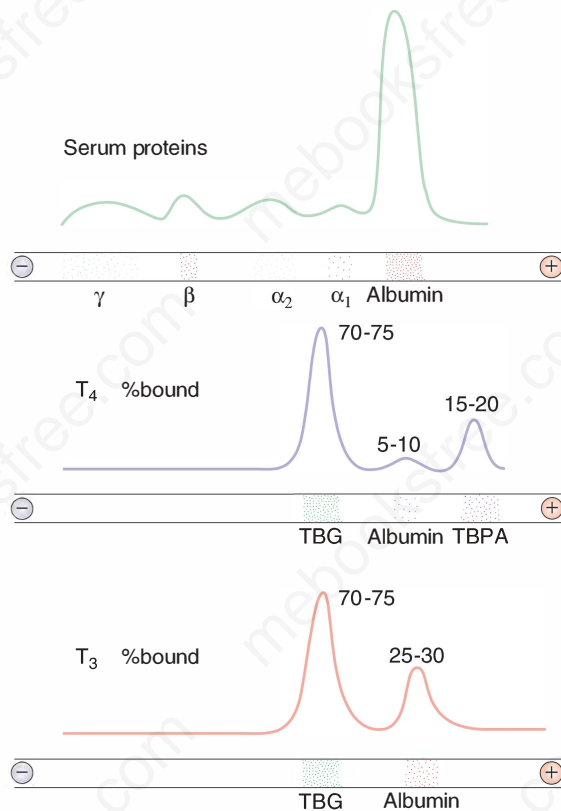
FIGURE 7–12 Diagrammatic representation of the distribution of radioactive T_4 and T_3 among serum thyroid hormone–binding proteins. **Top:** Paper electrophoretic pattern of serum proteins. **Middle:** Radioactive T_4 was added to serum and was then subjected to paper electrophoresis. The peaks represent the mobility of radioactive T_4 bound to different serum proteins. (TBG, thyroid hormone–binding globulin; TBPA, thyroxine–binding prealbumin; Transthyretin.) **Bottom:** Radioactive T_3 was added to serum and subjected to paper electrophoresis. The peaks indicate the relative distribution of protein-bound radioactive T_3. The figures above each peak indicate the relative hormone distribution among the binding proteins in a normal adult. (Reproduced with permission from James AE, Wagner HN, Cooke RE. *Pediatric Nuclear Medicine*. Philadephia: Saunders; 1974.)

TABLE 7–1 Causes of abnormal serum thyroxine determinations in euthyroid individuals.

Euthyroid Hyperthyroxinemia
↑ Plasma protein binding
↑ Thyroxine-binding globulin (TBG)
Inherited
Estrogen effect (pregnancy, estrogen therapy)
Hepatitis
Drugs: tamoxifen, 5-fluorouracil, clofibrate, methadone, heroin
↑ Transthyretin binding
Inherited
Paraneoplastic production by hepatic and pancreatic tumors
↑ Albumin binding
Inherited (familial dysalbuminemic hyperthyroxinemia)
↓ T_4-to-T_3 conversion[a]
Systemic illness
Medications: amiodarone, radiocontrast agents, glucocorticoids, propranolol
Thyroxine therapy in hypothyroidism[a]
Generalized resistance to thyroid hormone[a]
T_4 antibody
Biotin (assay interference)
Euthyroid Hypothyroxinemia
↓ Thyroxine-binding globulin
↓ TBG production
Inherited
Androgens
Drugs: danazol, L-asparaginase
↑ TBG clearance
Nephrotic syndrome
Severe liver disease
Protein-losing gastroenteropathy
Systemic illness[b]
Medications
Exogenous thyromimetic compounds (T_3 [Cytomel])[b]
Phenytoin and carbamazepine[b]
Iodine deficiency (with normal serum T_3)[b]

[a]Both total and free T_4 elevated.

[b]Both total and free T_4 low.

later) preserves normal free-hormone concentrations by lowering serum total thyroid hormone levels. Similarly, heparin stimulation of lipoprotein lipase releases free fatty acids that displace thyroid hormones from TBG. *In vivo*, this can result in lower total thyroid hormone concentrations, whereas *in vitro* (eg, in blood drawn through a heparin lock), heparin increases measured levels of free T_4 and T_3.

Transthyretin (Thyroxine-Binding Prealbumin)

Transthyretin, a 55-kD globular polypeptide composed of four identical 127-amino acid subunits, binds 10% of circulating T_4. Its affinity for T_4 is 10-fold greater than for T_3. The dissociation of T_4 and T_3 from transthyretin is rapid, so that transthyretin is a

source of readily available T_4. Increased affinity of transthyretin binding for T_4 can occur as a heritable condition. Affected individuals have an elevated total T_4 but a normal free T_4. Ectopic production of transthyretin, which has been reported to occur in patients with pancreatic and hepatic tumors, also causes euthyroid hyperthyroxinemia.

Albumin

Albumin binds to T_4 and T_3 with lesser affinity than TBG or transthyretin, but its high plasma concentration results in its transport of 15% of circulating T_4 and T_3. Rapid thyroid hormone dissociation rates from albumin make it an additional source of free hormone to tissues. Hypoalbuminemia, as occurs in nephrosis or cirrhosis, is associated with a low total T_4 and T_3, but the free-hormone levels are normal.

Familial dysalbuminemic hyperthyroxinemia is an autosomal dominant inherited disorder in which 25% of the albumin exhibits a higher than normal T_4-binding affinity. This results in an elevated total T_4 level but a normal free T_4 concentration and euthyroidism. In most affected kindreds, the T_3-binding affinity is

normal. Because these albumin variants do not bind the T_4 analogs used in many free T_4 immunoassays, these may falsely report elevation of the free T_4 in affected individuals.

METABOLISM OF THYROID HORMONES

The normal thyroid gland secretes about 100 nmol of T_4 and only 5 nmol of T_3 daily; less than 5 nmol of metabolically inactive reverse T_3 (rT_3) is produced (Figure 7–13). Most of the plasma pool of T_3 (80%) is derived from peripheral outer ring- or 5′-monodeiodination of T_4 in tissues outside of the thyroid gland, particularly the liver, kidney, and skeletal muscle (Table 7–2). Because T_3 has a higher binding affinity for the nuclear T_3 receptors that affect thyroid hormone action, this 5′-monodeiodination generates a more biologically active iodothyronine. On the other hand, deiodination of the *inner* ring of T_4 (5-deiodination) produces 3,3′,5′-triiodothyronine or reverse T_3 (rT_3), which is metabolically inert.

The three deiodinase enzymes that catalyze these reactions differ in tissue localization, substrate specificity, and physiologic and pathophysiologic modulation, as summarized in Table 7–2. Type 1 5′-deiodinase, the most abundant form, is found predominantly in liver and kidney, and in lesser quantities in the thyroid gland, skeletal and heart muscle, and other tissues. The type 1 5′-deiodinase contains a selenocysteine moiety that is probably at the active deiodinating site. The major function of type 1 5′-deiodinase is to provide T_3 to the circulation. Its activity is increased in hyperthyroidism and decreased in hypothyroidism. This partially accounts for the relatively higher circulating T_3 levels compared to T_4 levels seen in hyperthyroid patients. The enzyme is inhibited by the thionamide antithyroid drug PTU, but not methimazole, and by the antiarrhythmic drug amiodarone and iodinated radiocontrast dyes, such as sodium ipodate. Dietary selenium deficiency can also impair T_4 to T_3 conversion.

Type 2 5′-deiodinase is predominantly expressed in the brain and pituitary gland, where it maintains a constant level of intracellular T_3 in the central nervous system. Recent studies also have localized type 2 deiodinase to skeletal muscle, but the relative contribution of type 2 deiodinase derived T_3 to the total circulating pool of T_3 is controversial. This deiodinase is very sensitive to circulating T_4, so a lower circulating T_4 rapidly increases this enzyme's concentration in brain and
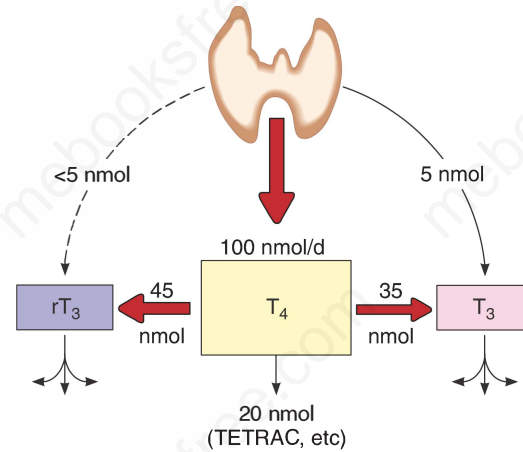
FIGURE 7–13 Major pathways of thyroxine metabolism in normal adult humans. Rates are expressed in nmol/24 h and are approximations based upon available data. 100 nmol of T_4 is equivalent to approximately 75 μg. (rT_3, reverse T_3; TETRAC, tetraiodothyroacetic acid.) (Reproduced with permission from Cavalieri RR, Rapoport B. Impaired peripheral conversion of thyroxine to triiodothyronine. *Annu Rev Med.* 1977;28:57-65.)

pituitary—maintaining the intracellular T_3 level and sustaining its neuronal cellular functions. Conversely, elevated plasma T_4 reduces the type 2 5′-deiodinase level, protecting brain cells from excessive T_3. Consequently, this deiodinase represents a mechanism by which the hypothalamus and pituitary can be buffered to some extent from changes in the circulating T_4 level. rT_3 can also modify the activity of type 2 5′-deiodinase in the brain and the pituitary gland; and α-adrenergic compounds stimulate type 2 5′-deiodinase in brown fat, but the physiologic significance of these effects is not clear.

Type 3 5′-deiodinase is found in chorionic membranes of the placenta and glial cells in the central nervous system, where it inactivates T_4 by converting it to rT_3 and it inactivates T_3 by converting it to 3,3′-diiodothyronine (3,3′-T_2) (Figure 7–14). Type 3 deiodinase levels are elevated in hyperthyroidism and decreased in hypothyroidism and, therefore, may help insulate the fetus and the brain from T_4 excess or deficiency. Placental type 3 deiodinase accelerates T_4 disposal in pregnant women, which partly explains the increased T_4 dose requirements in treated hypothyroid women. Congenital and acquired hemangiomas can express high levels of the type 3 deiodinase and cause a rare form of consumptive hypothyroidism.

TABLE 7–2 Iodothyronine deiodinase types and characteristics.

Deiodinase Type	D1	D2	D3
Substrates	$rT_3 > T_4 > T_3$	$T_4 > rT_3$	$T_3 > T_4$
Tissue distribution	Liver, kidney, thyroid	Brain, pituitary, skeletal muscle	Brain, placenta, fetal tissues
Function	Plasma T_3 production	Local and plasma T_3 production	T_3 degradation
PTU inhibition (IC50, μM)	5	>1000	>1000
Hypothyroidism	Decrease	Increase	Decrease
Hyperthyroidism	Increase	Decrease	Increase
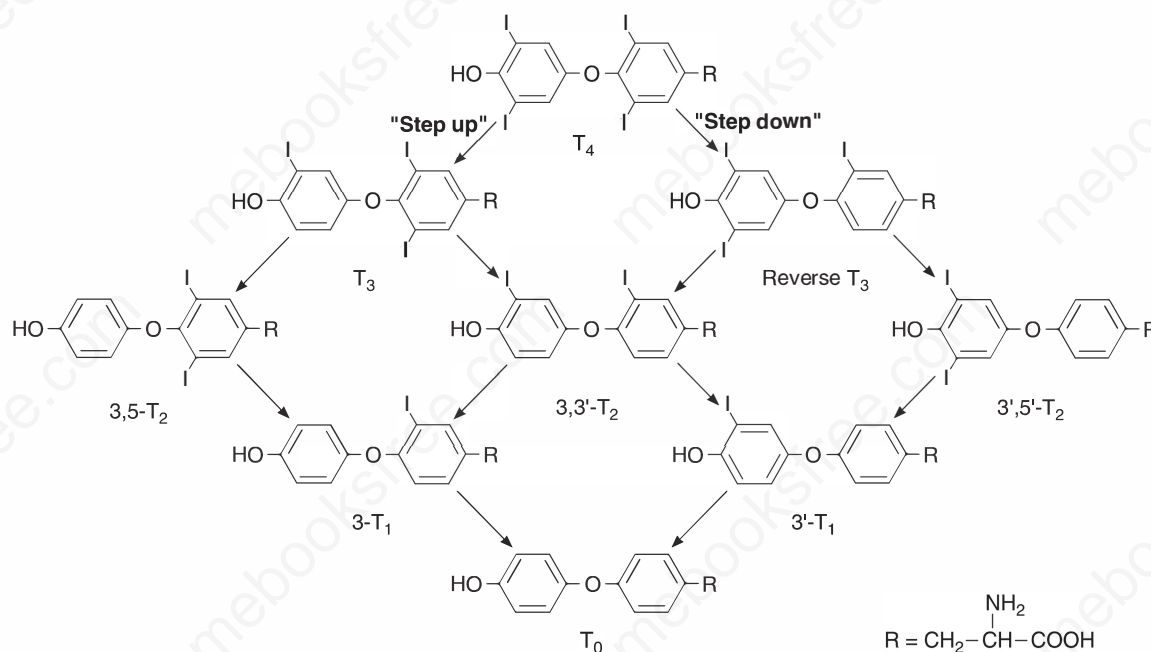
FIGURE 7–14 The deiodinative pathway of thyroxine metabolism. The monodeiodination of T_4 to T_3 represents a step-up in biologic potency, whereas the monodeiodination of T_4 to reverse T_3 has the opposite effect. Further deiodination of T_3 essentially abolishes hormonal activity.

Overall, the functions of the deiodinases are physiologically important in three ways. First, they permit local tissue and cellular modulation of thyroid hormone actions. Second, they help the organism adapt to changing states, including iodine deficiency or chronic illness. Third, they regulate thyroid hormone actions in the early development of many vertebrates, including amphibia and mammals.

About 80% of T_4 is metabolized by deiodination, 35% to T_3 and 45% to rT_3 (see Figure 7–13). The remainder is inactivated mainly by glucuronidation in the liver and biliary secretion and, to a lesser extent, by sulfation in the liver or kidney. Other metabolic reactions include deamination of the alanine side chain, forming thyroacetic acid derivatives of low-biologic activity; or decarboxylation or cleavage of the ether link, forming inactive compounds.

As a result of these metabolic pathways, approximately 10% of the total 1000 nmol extrathyroidal T_4 pool is cleared each day, and the plasma half-life of T_4 is 7 days. Due to the lower binding affinity of plasma proteins for T_3, the turnover of its smaller extrathyroidal pool is more rapid, with a plasma half-life of 1 day. The total body pool of rT_3 is about the same size as that of T_3, but rT_3 has an even more rapid turnover, with a plasma half-life of only 0.2 day.

CONTROL OF THYROID FUNCTION AND HORMONE ACTION

Growth and function of the thyroid gland are controlled by the hypothalamic-pituitary-thyroid axis (Figure 7–15) and, as previously discussed, by iodide through the elements of autoregulation.
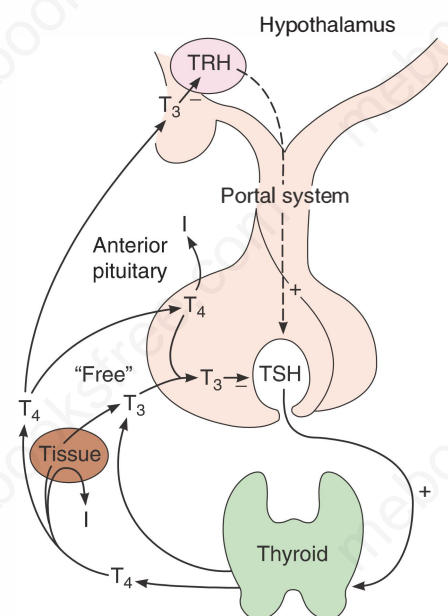
FIGURE 7–15 The hypothalamic-hypophysial-thyroidal axis. TRH produced in the hypothalamus reaches the thyrotrophs in the anterior pituitary by the hypothalamo-hypophysial portal system and stimulates the synthesis and release of TSH. In both the hypothalamus and the pituitary, it is primarily T_3 that inhibits TRH and TSH secretion, respectively. T_4 undergoes monodeiodination to T_3 in neural and pituitary as well as in peripheral tissues.

Hypothalamic thyrotropin-releasing hormone (TRH) stimulates thyrotrophic cells in the anterior pituitary to produce TSH, which in turn promotes thyroid gland growth and hormone secretion. In addition, deiodinases in the pituitary and peripheral tissues modulate thyroid hormone effects by their tissue-specific conversion of T_4 to the more active iodothyronine T_3. Finally, the molecular effects of T_3 in individual tissues are modulated by the subtype of T_3 receptor with which it interacts; the specific gene activation or repression response that it induces; and in a realm just now being revealed, the T_3 receptor's interaction with other ligands, closely related receptors (eg, retinoid X receptor; RXR), and coactivators and corepressors that interact with it in modulating gene expression. Mutations in the T_3 receptor isoforms cause syndromes of thyroid hormone resistance (see section Resistance to Thyroid Hormone [RTH] Syndromes).

Thyrotropin-Releasing Hormone

TRH is a tripeptide, pyroglutamyl-histidyl-proline amide (pyro-Glu-His-Pro-NH_2), synthesized by neurons in the supraoptic and supraventricular nuclei of the hypothalamus (Figure 7–16). TRH is stored in the median eminence of the hypothalamus and then transported via the pituitary portal venous system down the pituitary stalk to the anterior pituitary, where it controls synthesis and release of TSH. TRH is also found in other portions of the hypothalamus, brain, and spinal cord, where it may have distinct functions as a neurotransmitter. The pre-pro-*TRH* gene, which encodes a large molecule with five copies of the TRH progenitor sequence Glu-His-Pro-Gly, is located on chromosome 3. *TRH* gene expression is negatively regulated by thyroid hormone—both the T_3 delivered by the circulation and that arising from T_4 deiodination in peptidergic neurons themselves (see Table 7–2).

In the anterior pituitary, TRH binds to a specific membrane receptor located on TSH- and prolactin-secreting cells, stimulating synthesis and release of their respective hormones. The TRH receptor is a member of the 7-transmembrane-spanning, G protein–coupled receptor family (Table 1–1; Figure 1–4). TRH binds to the receptor's third transmembrane helix, activating both its cyclic guanosine monophosphate-producing complex and the inositol 1,4,5-triphosphate (IP_3) signaling cascade that releases intracellular Ca^{2+} and generates 1,2-diacylglycerol, thereby activating protein kinase C. These pathways are responsible for stimulating TSH release. They coordinate transcription of the genes encoding the TSH subunits and posttranslational glycosylation of TSH, which is necessary for full biologic activity.

(pyro)Glu-His-Pro-(NH₂)
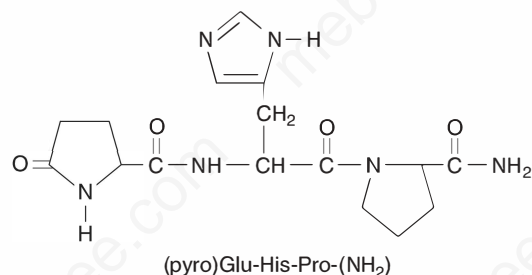
FIGURE 7–16 Chemical structure of thyrotropin-releasing hormone (TRH).

TRH-stimulated TSH secretion is pulsatile (Figure 7–17), with a mean TSH pulse amplitude of 0.6 mU/L every 2 hours. Normal individuals have a circadian rhythm in TSH release, with a peak circulating level between midnight and 4 AM, which is presumably controlled by a hypothalamic neuronal pulse generator driving TRH synthesis.

Thyroid hormones exert additional negative feedback on TSH production at the level of the pituitary by downregulating the number of TRH receptors on pituitary thyrotropes. Consequently, in patients with hyperthyroidism, both TSH pulses and its nocturnal surge are markedly suppressed; whereas in hypothyroid patients, the TSH pulse amplitude and nocturnal surge are much greater. In experimental animals and newborn humans, exposure to cold temperature increases TRH and TSH secretion. Certain other hormones and drugs (eg, vasopressin and α-adrenergic agonists), stimulate TRH synthesis and release. Leptin also modulates TRH and TSH secretion and may be responsible for the reversible TSH rise seen in some obese patients.

When synthetic TRH is administered intravenously to humans as a 200 to 500 μg bolus, it generates a prompt three- to fivefold rise in the serum TSH concentration, peaking at about 30 minutes and lasting for 2 to 3 hours (see Figure 4–14). In patients with primary hypothyroidism, in whom basal TSH is elevated, there is an exaggerated TSH response to exogenous TRH; and this response is suppressed in patients with hyperthyroidism, high-dose T_4 therapy, and central hypothyroidism.

TRH is also found in the islet cells of the pancreas, gastrointestinal tract, placenta, heart, prostate, testes, and ovaries. TRH production in these peripheral tissues is not inhibited by T_3; the role of TRH in these tissues remains unknown.

Thyrotropin (Thyroid-Stimulating Hormone)

TSH is 28-kD glycoprotein composed of α and β subunits that are noncovalently linked. The α subunit is common to the two other pituitary glycoproteins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), and the placental hormone human chorionic gonadotropin (hCG), whereas the β subunit is unique for each glycoprotein hormone, conferring specific binding properties and biologic activity. The genes for the TSH α and β subunits are located on chromosomes 6 and 1, respectively. The human α subunit has an apoprotein core of 92 amino acids and contains two oligosaccharide chains; the TSH β subunit has an apoprotein core of 112 amino acids and contains one oligosaccharide chain. The α and β subunit amino acid chains of TSH each form three loops that are intertwined into a cystine knot (Figure 7–18). Glycosylation takes place in the RER and the Golgi, where glucose, mannose, and fucose residues and terminal sulfate or sialic acid residues are linked to the apoprotein core. These carbohydrate residues prolong its plasma half-life and enhance it ability to induce TSH receptor activation.

TSH controls thyroid cell growth and hormone production by binding to a specific TSH receptor, one of approximately 1000 located on the basolateral cell membrane of each thyroid cell. TSH binding activates both the cyclic adenosine monophosphate
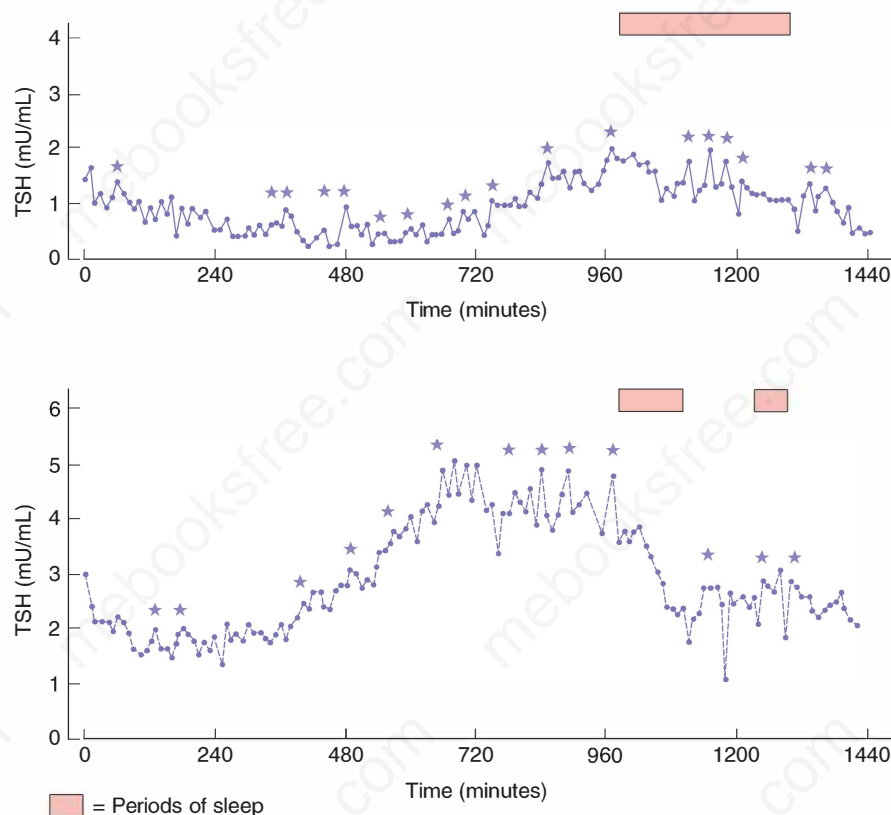
FIGURE 7–17 Serum TSH in two normal subjects demonstrating spontaneous pulses and the circadian rhythm of TSH secretion. (0 time is 0900; stars indicate significant pulses.) (Reproduced with permission from Greenspan SL, Klibanski A, Schoenfeld D, et al. Pulsatile secretion of thyrotropin in man. *J Clin Endocrinol Metab*. 1986 Sep;63(3):661-668.)

(cAMP) and the phosphoinositol pathways for signal transduction. The TSH receptor gene is located on chromosome 14; the product is a single-chain glycoprotein composed of 764 amino acids. The TSH receptor is a member of the 7-membrane spanning, G protein–coupled receptor family, with an ectodomain involved in ligand binding and an intramembrane and intracellular portion responsible for activation of the signaling pathways that promote thyroid cell growth and hormone synthesis and release (Figure 7–19).

The TSH receptor is involved in the pathogenesis of numerous congenital and acquired forms of hypothyroidism and hyperthyroidism. Heritable defects resulting in impaired TSH synthesis or action have been described, including mutant genes for transcription factors required for pituitary thyrotrope differentiation (*POU1F1, PROP1, LHX3, HESX1*), TRH receptor, TSH β-chain, TSH receptor, and Gsα, which transduces TSH receptor binding to adenylate cyclase activation. Acquired TSH receptor blocking antibodies can also cause hypothyroidism.

The most common TSH receptor–related disorder causing hyperthyroidism is Graves disease, in which autoantibodies bind and stimulate the TSH receptor. However, the TSH receptor is involved in the etiology of several other forms of hyperthyroidism. Germline mutations activating the TSH receptor can cause familial hyperthyroidism, and somatic activating mutations result in toxic adenomas. Other mutations can lead to aberrant TSH receptor activation by hCG, the placental glycoprotein hormone that is structurally similar to TSH, in familial gestational hyperthyroidism.

Effects of TSH on the Thyroid Cell

TSH has many actions on the thyroid cell. Most of its actions are mediated through the G protein-adenylyl cyclase-cAMP system, but activation of the phosphatidylinositol (PI) system with a resulting increase in intracellular calcium is also involved. The major actions of TSH include the following:

A. Changes in thyroid cell morphology TSH rapidly induces pseudopods at the follicular cell–colloid border, accelerating TG resorption. Colloid content is diminished as intracellular colloid droplets are formed, and lysosome formation is stimulated, increasing TG hydrolysis and thyroid hormone release.

B. Cell growth Individual thyroid cells increase in size; vascularity is increased; and, over a period of time, thyroid enlargement, or goiter, develops.

C. Iodine metabolism TSH stimulates all phases of iodide metabolism, from increased iodide uptake and transport to increased iodination of TG and increased secretion of thyroid hormones and TG itself. Increased NIS expression and
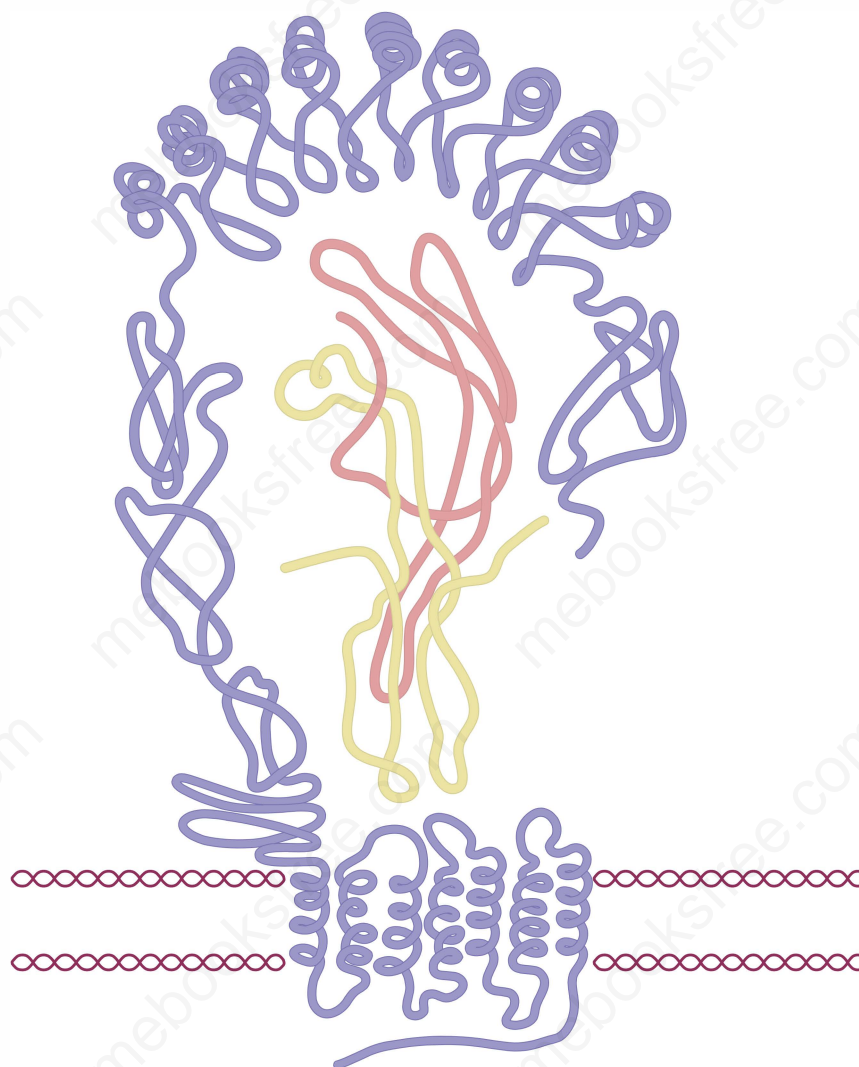
FIGURE 7–18 Schematic configuration of the TSH-TSHR complex. The central portion of the figure represents the ribbon-like structure of TSH within the TSH receptor. The pink line represents the β subunit and the yellow line the α subunit. See also Figure 7–19. (Data from Kajava AV, Vassart G, Wodak SJ. Modeling of the three-dimensional structure of proteins with the typical leucine-rich repeats. *Structure*. 1995 Sep 15;3(9):867-877.)

the stimulation of cAMP production mediate increased iodide transport, and phosphatidylinositol-4,5-bisphosphate (PIP_2) hydrolysis and increased intracellular Ca^{2+} stimulate the iodination of TG. The TSH effect on thyrocyte iodide content is biphasic: initially, it is decreased (iodide efflux) due to the rapid increase in hydrolysis of TG with release of hormone and leakage of iodide out of the gland; and then after a lag of several hours, iodide uptake is increased. The TSH-mediated stimulation of thyrocyte iodine uptake and TG secretion also occur after administration of recombinant TSH for radioiodine treatment and monitoring of patients with well-differentiated thyroid cancers.

D. Other effects of TSH   Other effects include increased transcription of the mRNAs for TG and TPO; increased incorporation of iodide into MIT, DIT, T_3, and T_4; and increased

lysosomal activity, with increased secretion of T_4 and T_3 from the gland. There is also increased activity of type 1 5′-deiodinase, which helps conserve intrathyroidal iodine.

TSH has additional effects on the thyroid gland, including stimulation of glucose uptake, oxygen consumption, and glucose oxidation—effects that can enhance detection of thyroid cancer metastases by Fluorodeoxyglucose Positron Emission Tomography (FDG PET) scanning. There is accelerated turnover of phospholipids and stimulation of synthesis of purine and pyrimidine precursors, with increased synthesis of DNA and RNA.

Serum TSH

Intact TSH and isolated α subunit are both present in circulating blood and detectable by immunoassay in concentrations that are
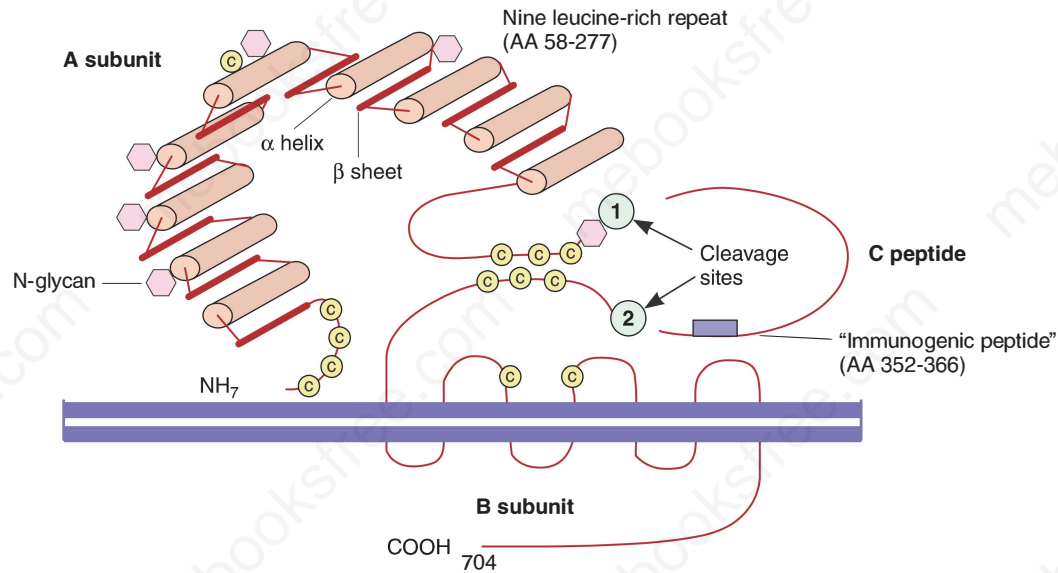
FIGURE 7–19 Schematic representation of the TSH receptor. The A subunit is the ligand-binding portion of the receptor and the B subunit is the activation portion. The ligands which bind to the receptor include TSH, TSH-stimulating antibody, and TSH-blocking antibody. There are two cleavage sites which allow breakage of the receptor and loss of the A subunit into the serum. (Reproduced with permission from Rapoport B, Chazenbalk GD, Jaume JC, et al. The thyrotropin (TSH) receptor: interaction with TSH and autoantibodies. *Endocr Rev*. 1998 Dec;19(6):673-716.)

normally 0.5 to 4.0 mU/L and 0.5 to 2 μg/L, respectively. The serum TSH level is increased in primary hypothyroidism and decreased in thyrotoxicosis, whether endogenous or from excessive oral intake of thyroid hormones. The plasma half-life of TSH is about 30 minutes, and the daily production rate is about 40 to 150 mU/d.

The glycoprotein α subunit is often disproportionately elevated in patients with TSH-secreting pituitary tumors (see later); it is also increased in normal postmenopausal women due to increased synthesis and secretion related to increased gonadotropin production.

Control of Pituitary TSH Secretion

Two major factors control synthesis and release of TSH: (1) T_3 level within thyrotroph cells, which regulates mRNA expression, TSH translation, and the hormone's release; and (2) TRH, which controls posttranslational glycosylation and also release.

TSH synthesis and release are inhibited by high serum levels of T_4 and T_3 (thyrotoxicosis) and stimulated by low levels of thyroid hormone (hypothyroidism). Thyrotoxicosis can suppress serum TSH levels beneath the limits of assay detection, and recovery of normal TSH secretion may require weeks or months after restoration of normal thyroid hormone levels. Certain other hormones and drugs can also inhibit TSH secretion. These include somatostatin, dopamine, dopamine agonists such as bromocriptine, and high doses of glucocorticoids—all of which typically lower the circulating TSH level, but do not suppress it completely. Severe systemic illnesses can inhibit TSH secretion, and there may be a rebound increase in TSH as the patient recovers. In contrast, overt hyperthyroidism can suppress the TSH concentration to beneath the detection limits of even the most sensitive TSH immunoassays.

Tumors and other disorders of the hypothalamus or anterior pituitary gland can impair TRH and TSH secretion, respectively. Pituitary disorders causing hypothyroidism are termed *secondary*, whereas hypothalamic disease resulting in hypothyroidism is called *tertiary*. Differential diagnosis of these lesions is discussed later (see Thyroid Tests).

Other Thyroid Stimulators and Inhibitors

Growth factors such as insulin, insulin-like growth factor-1, insulin-like growth factor binding proteins, epidermal growth factor, transforming growth factor β1, and fibroblast growth factors and their receptors, as well as autocrine factors such as prostaglandins and cytokines, modify thyroid cell growth and hormone production. Some of these may be involved in the pathogenesis of benign nodular goiter. The thyroid follicle has a rich supply of capillaries that are accompanied by noradrenergic nerve fibers from the superior cervical ganglion and acetylcholine esterase-positive nerve fibers derived from the vagus nerve and thyroid ganglia. However, these are not known to play any role in thyroid gland regulation. The parafollicular C cells secrete both calcitonin and calcitonin gene-related peptide (CGRP). In experimental animals, these and other neuropeptides modify thyroid blood flow and hormone secretion.

The Actions of Thyroid Hormones
1. THE THYROID HORMONE RECEPTORS AND MECHANISMS OF ACTION

The thyroid hormones exert their actions through two general mechanisms: (1) genomic actions effected through T_3 interactions with its nuclear receptors, regulating gene activity; and

		T₃ binding	DNA binding
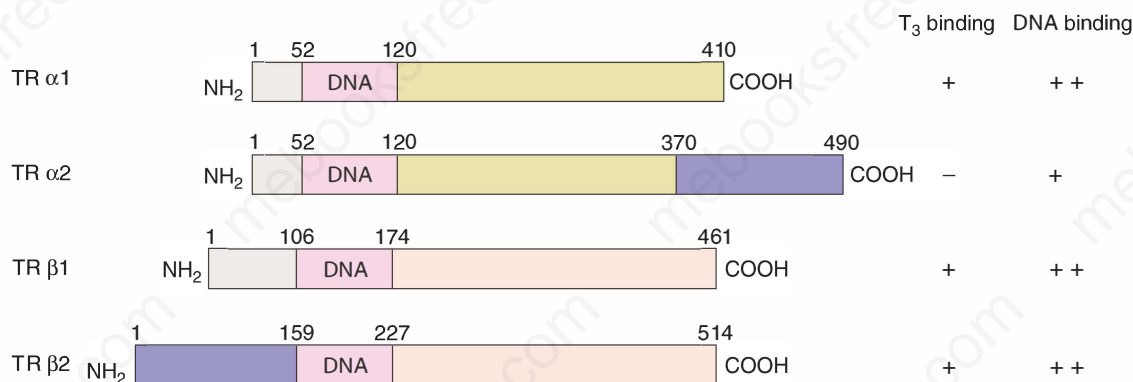
FIGURE 7–20 Deduced protein structure of the thyroid hormone receptor α and β gene products. The receptor protein has three domains: a DNA-binding domain with a high degree of similarity among the different types of receptors, a carboxyl terminal triiodothyronine (T_3)–binding domain, and an amino terminal domain that is not required for full function. The numbers above the structures represent amino acid numbers. The properties of the receptors with respect to their ability to bind T_3 and bind to a T_3-response element of DNA are shown on the right. Identical shading of receptor domains indicates identical amino acid sequences. (TR, thyroid hormone receptor.) (Reproduced with permission from Brent GA. The molecular basis of thyroid hormone action. *N Engl J Med*. 1994 Sep 29;331(13):847-853.)

(2) nongenomic actions mediated by T_3 and T_4 interactions with certain enzymes (eg, calcium ATPase, adenylate cyclase, and monomeric pyruvate kinase), glucose transporters, mitochondrial, and membrane proteins.

Thyroid hormones, that are unbound in plasma, are transported intracellularly, by either specific carriers including monocarboxylate transporter 8 (MCT8), MCT10, and organic anion transporting polypeptide (OATP1C1). OATP1C1 is expressed predominantly in brain capillaries and the choroid plexus, and transports T_4 preferentially, while MCT8 and MCT10 are expressed in many tissues and transport both T_4 and T_3. Thyroid hormones are transported through the cell membrane into the cytoplasm, and subsequently into the nucleus, where T_3 binds to its specific receptor. Clinical syndromes associated with mutations in these transporters have recently been recognized, such as the Allan-Herndon-Dudley syndrome in which affected male children develop mental retardation, myopathy, and a movement disorder in association with elevated serum T_3 and low serum T_4 levels.

The T_3 nuclear receptor is member of a nuclear receptor superfamily that includes nuclear receptors for glucocorticoids, mineralocorticoids, estrogens, progestins, vitamin D_3, and retinoids (see Figure 1–16). There are two thyroid hormone receptor (*TR*) genes in humans: *TR*α located on chromosome 17 and *TR*β on chromosome 3. Each gene yields at least two differentially spliced products, TRα 1 and 2 and TRβ 1 and 2, although TRα 2 is believed to be biologically inactive (Figure 7–20). Each of these receptors has a carboxyl terminal ligand-binding domain and a centrally located DNA-binding domain with two cysteine zinc fingers that facilitate their specific attachment to thyroid hormone response elements (TREs) in the promoters of target genes and regulate their transcription (see Figures 1–16 and 1–17). The concentration of these receptors in tissue varies among tissues and with their stage of development. For example, the brain contains predominantly TRα, the liver mostly TRβ, and cardiac muscle contains both. Point mutations in the ligand-binding domain of the *TR*β gene have been shown to be responsible for the syndrome of generalized resistance to thyroid hormone (GRTH). The TRs bind to TREs, which are typically paired, specific oligonucleotide sequences (eg, AGGTCA) (Figure 7–21). The TRs can also function as heterodimers with receptors for other transcription factors, such as the retinoid X and retinoic acid receptors. TREs are generally located upstream of the transcription start site for the coding regions of thyroid hormone–responsive genes. In positively regulated genes, unbound TRs interact with corepressors (eg, nuclear receptor corepressor [NCoR] and silencing mediator for retinoic and thyroid hormone receptors [SMRT]) to repress basal transcription by recruiting histone deacetylases that alter the nearby chromatin structure. When TRs are bound by T_3, these corepressor complexes are released, and the T_3-bound TRs associate with coactivator complexes that promote local histone acetylation. They also associate with another protein complex (vitamin D receptor-interacting protein/TR-associated proteins) that recruits RNA polymerase II and starts gene transcription. Some genes are negatively regulated by T_3-bound TRs, such as pre-pro-*TRH* and *TSH*-α and -β subunit genes. Thyroid hormone's actions to alter expression levels of specific mRNAs and their translated proteins generate a constellation of specific tissue responses (see Figure 7–21).

Additional nongenomic actions of thyroid hormones have long been suspected based on certain rapid tissue responses that take place before RNA transcription could occur and by recognition of T_3- and T_4-binding sites outside of the nucleus. For example, it has been shown that thyroid hormones bind the membrane protein integrin αVβ3, which is involved in thyroid hormone–mediated activation of the MAP kinase cascade and stimulation of angiogenesis.

2. PHYSIOLOGIC EFFECTS OF THYROID HORMONES

The transcriptional effects of T_3 characteristically demonstrate a lag time of hours or days to achieve full effect. These genomic actions have a number of vital effects, including tissue growth,
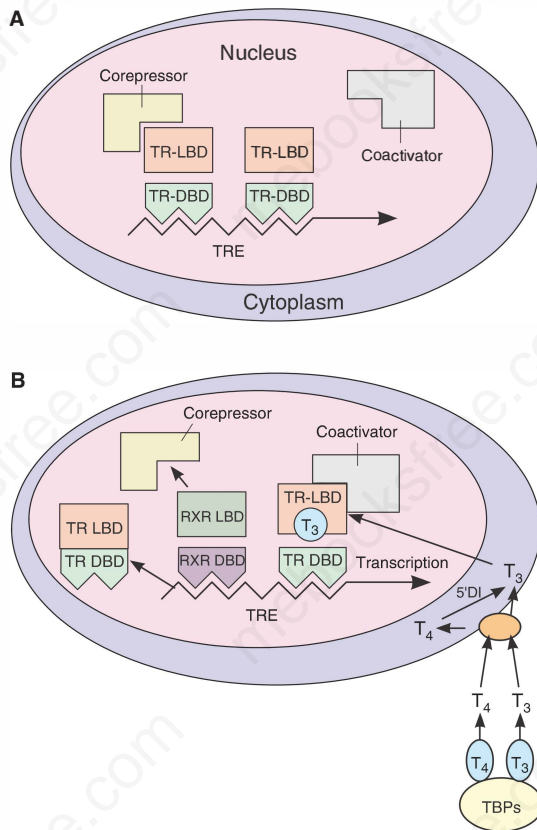
FIGURE 7–21 Model of the interaction of T_3 with the T_3 receptor. **Panel A, Inactive Phase:** The unliganded T_3 receptor dimer bound to the TRE along with corepressors acts as a suppressor of gene transcription. **Panel B, Active Phase:** T_3 and T_4 circulate bound to thyroid-binding proteins (TBPs). The free hormones are transported into the cell by a specific transport system. Within the cytoplasm, T_4 is converted to T_3 by 5′-deiodinase and T_3 moves into the nucleus. There it binds to the ligand-binding domain of the TR monomer. This promotes disruption of the TR homodimer and heterodimerization with RXR on the TRE, displacement of corepressors, and binding of coactivators. The TR-coactivator complex activates gene transcription, which leads to alteration in protein synthesis and cellular phenotype. (TR-LBD, T_3 receptor ligand-binding domain; TRDBD, T_3 receptor DNA-binding domain; RXR-LBD, retinoid X receptor ligand-binding domain; RXR-DBD, retinoid X receptor DNA-binding domain; TRE, thyroid hormone response element; TBPs, thyroxine-binding proteins; T_3, triiodothyronine; T_4, tetraiodothyronine, L-thyroxine; 5′DI, 5′-deiodinase.)

brain maturation, increased calorigenesis and oxygen consumption, as well as other specific effects on the heart, liver, kidneys, skeletal muscle, and skin. However, as noted earlier, some actions of T_3 are believed not to be genomic, including its reduction of pituitary type 2 5′-deiodinase activity and the increased glucose and amino acid transport that it can induce in some tissues. Some specific effects of thyroid hormones are summarized later.

Effects on Fetal Development

Iodide concentration by thyroid tissue and pituitary TSH both appear in the human fetus at about 11 weeks of gestation. Because

of the high placental content of type 3 5′-deiodinase, most maternal T_3 and T_4 are inactivated, and very little free hormone reaches the fetal circulation. However, this small amount of free hormone from the mother may be important for early fetal brain development. After 15 to 18 weeks of gestation, the fetus is largely dependent on its own thyroidal secretion. Although some fetal growth occurs in the absence of fetal thyroid hormone secretion, brain development and skeletal maturation are markedly impaired if congenital hypothyroidism is undiagnosed and thyroid hormone therapy is not begun promptly after birth. Congenital hypothyroidism results in cretinism, the features of which include mental retardation and dwarfism.

Effects on Oxygen Consumption, Heat Production, and Free Radical Formation

T_3 increases O_2 consumption and heat production in part by stimulation of Na$^+$-K$^+$ ATPase in all tissues except the brain, spleen, and testis. This contributes to the increased basal metabolic rate (total somatic O_2 consumption at rest) and the increased sensitivity to heat in hyperthyroidism—and the opposite in hypothyroidism. Thyroid hormones stimulate mitochondriogenesis, augmenting the cell's oxidative capacity. They also induce changes in the mitochondrial inner membrane protein and lipid composition that increase oxidative metabolism by both genomic and nongenomic effects. The reduced efficiency of oxidative metabolism caused by thyroid hormone is also reflected in the increased futile cycling of intermediary carbohydrate metabolites.

Cardiovascular Effects (Figure 7–22)

T_3 stimulates transcription of sarcoplasmic reticulum Ca^{2+} ATPase, increasing the rate of myocardial diastolic relaxation. It also increases expression of the more rapidly contractile α isoform of myosin heavy chain, which contributes to enhanced systolic function. In myocardium, T_3 also alters expression of different isoforms of the Na$^+$-K$^+$ ATPase genes, increases expression of α-adrenergic receptors, and decreases the concentration of the inhibitory G protein Gi α. Thyroid hormones also lower peripheral vascular resistance and increase intravascular volume, which contribute further to the increase in cardiac output associated with thyroid hormone action. T_3 also increases the rates of both depolarization and repolarization of the sinoatrial node, increasing heart rate. Consequently, thyroid hormones have positive chronotropic and inotropic effects on the heart, which, along with the heightened adrenergic sensitivity (see later), account for the increased heart contractility and rate in hyperthyroidism and the reverse in hypothyroidism.

Sympathetic Effects

Thyroid hormones increase the number of β-adrenergic receptors in heart and skeletal muscle, adipose tissue, and lymphocytes.
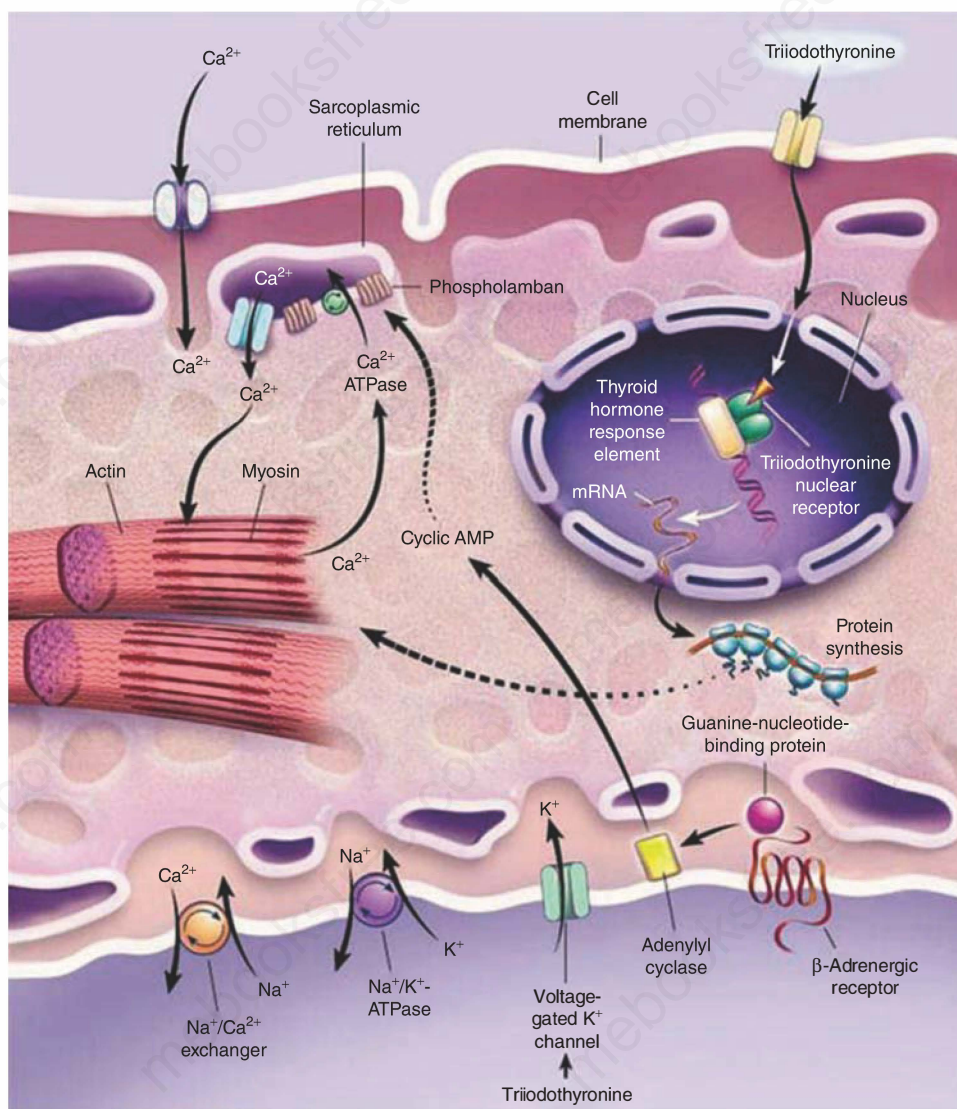
FIGURE 7–22 Sites of action of triiodothyronine on cardiac myocytes. Triiodothyronine enters the cell, possibly by a specific transport mechanism, and binds to nuclear triiodothyronine receptors. The complex then binds to thyroid hormone response elements of target genes and regulates transcription of these genes. Nonnuclear thyroid hormone effects on ion channels also occur. (Reproduced with permission from Klein I, Ojamaa K. Thyroid hormone and the cardiovascular system. *N Engl J Med.* 2001 Feb 15;344(7):501-509.)

They may also amplify catecholamine action at a postreceptor site. Many of the clinical manifestations of thyrotoxicosis appear to reflect increased sensitivity to catecholamines. Furthermore, therapy with β-adrenergic–blocking agents is often helpful in controlling these sympathomimetic manifestations of thyroid hormone excess.

Pulmonary Effects

Thyroid hormones maintain ventilatory responsiveness to hypoxia and hypercapnia in the brain stem respiratory center. Consequently, in patients with severe hypothyroidism, hypoventilation can occur. Respiratory muscle functions are also regulated by thyroid hormone, and can be weakened in hyperthyroidism, leading to a sense of breathlessness.

Hematopoietic Effects

The increased cellular demand for O_2 in hyperthyroidism leads to increased production of erythropoietin and increased erythropoiesis. However, blood volume is usually not increased because of hemodilution and increased red cell turnover. Thyroid hormones increase the 2,3-diphosphoglycerate content of erythrocytes, allowing increased O_2 dissociation from hemoglobin and increasing O_2 availability to tissues. The reverse occurs in hypothyroidism.

Gastrointestinal Effects

Thyroid hormones promote gut motility, which can result in increased motility and hyperdefecation (ie, increased frequency of formed bowel movements) in hyperthyroidism. Conversely, slowed bowel transit and constipation occur in hypothyroidism.

Skeletal Effects

Thyroid hormones stimulate bone turnover, increasing bone resorption and, to a lesser degree, bone formation. Consequently, hyperthyroidism can be associated with hypercalciuria, and less often, hypercalcemia. Furthermore, chronic thyroid hormone excess can cause clinically significant bone mineral loss and increased fracture risk.

Neuromuscular Effects

In hyperthyroidism, there is increased protein turnover and loss in skeletal muscle, which can lead to a characteristic proximal myopathy. There is also an increase in the speed of muscle contraction and relaxation, noted clinically as the hyperreflexia of hyperthyroidism, and the delayed deep tendon reflex relaxation phase in hypothyroidism. A fine distal hand tremor is also typical in hyperthyroidism. As noted earlier, thyroid hormones are essential for normal development and function of the central nervous system, and failure of fetal thyroid function results in severe mental retardation. In the adult, hyperactivity in hyperthyroidism and sluggishness in hypothyroidism can be striking. Similarly, hyperthyroid patients can be anxious or agitated, while hypothyroidism is typically associated with a depressed mood.

Effects on Lipid and Carbohydrate Metabolism

Hyperthyroidism increases hepatic gluconeogenesis and glycogenolysis, as well as intestinal glucose absorption, and there may also be thyroid hormone–mediated decreases in insulin sensitivity. Thus, hyperthyroidism can worsen glycemic control in patients who also have diabetes mellitus. Cholesterol synthesis and degradation are both increased by thyroid hormones. The latter effect is due largely to an increase in hepatic low-density lipoprotein (LDL) receptor number, accelerating LDL clearance. Consequently, total and LDL cholesterol levels are typically elevated in patients with hypothyroidism. In hyperthyroidism, lipolysis is increased, releasing fatty acids and glycerol into circulating plasma.

Endocrine Effects

Thyroid hormones alter the production, responsiveness, and metabolic clearance of a number of hormones. In hypothyroid children, impaired growth hormone release slows longitudinal growth. Hypothyroidism can cause delayed puberty by impairing gonadotropin-releasing hormone (GnRH) and gonadotropin secretion. Conversely, primary hypothyroidism can also cause precocious puberty, perhaps as an effect of very high TSH levels on gonadotropin receptors. In adults, hypothyroidism causes hyperprolactinemia in a minority of affected women. Menorrhagia (prolonged and heavier menses) and anovulation are common in hypothyroid women, the latter resulting in infertility. The responsiveness of the hypothalamic-pituitary-adrenal axis to stress is blunted in hypothyroid patients. A slowing of the cortisol metabolic clearance rate compensates for this in the hypothyroid state. Conversely, however, restoration of euthyroidism can rarely provoke adrenal insufficiency as cortisol metabolism is accelerated in patients with diminished cortisol reserve due to concomitant disease affecting the adrenal axis.

In hyperthyroidism, accelerated aromatization of androgens to estrogens and increased sex hormone–binding globulin levels contribute to the gynecomastia and elevated total testosterone levels seen in affected men. Hyperthyroidism can also impair normal GnRH and gonadotropin regulation of ovulation and menses, causing infertility and amenorrhea, respectively. All of the endocrine derangements occurring with thyroid dysfunction are reversed by appropriate treatment to restore euthyroidism.

PHYSIOLOGIC CHANGES IN THYROID FUNCTION

Thyroid Function in the Fetus

Before the fetus develops its own independent thyroid function, it depends on maternal thyroid hormones for early neural development. By the 11th week of gestation, however, the hypophysial portal system has developed, and measurable TSH and TRH are present. At about the same time, the fetal thyroid begins to trap iodine. The secretion of thyroid hormone probably begins in mid gestation (18-20 weeks). TSH increases rapidly to peak levels at 24 to 28 weeks, and T_4 levels peak at 35 to 40 weeks. T_3 levels remain low during gestation; T_4 is converted to rT_3 by type 3 5′-deiodinase during fetal development. At birth, there is a sudden marked rise in TSH stimulated by exposure to the colder extrauterine environment, with a rise in T_4, a rise in T_3, and a fall in rT_3. These parameters gradually stabilize over the first month of life.

Thyroid Function in Pregnancy

A number of thyroid parameters change during pregnancy. There is an increase in urinary iodide clearance; in areas of low dietary iodine intake, this can lead to maternal goiter, or when severe, hypothyroidism. During pregnancy, there is a rise in TBG due to an estrogen-induced increase in hepatic glycosylation of TBG, prolonging its metabolic clearance rate. Consequently, serum total T_4, and to a lesser extent, total T_3 concentrations rise. Rising levels of hCG, which has weak TSH receptor agonist activity, contribute to minimal thyroid enlargement. Maternal hCG levels peak at approximately 12 weeks resulting in a transient high-normal or even modestly elevated serum-free T_4 level and a physiologic suppression of serum TSH to the low-normal or even subnormal range. Pathologically elevated hCG levels, found in women with a molar pregnancy or choriocarcinoma, can cause overt hyperthyroidism. Women with severe vomiting in pregnancy, called hyperemesis gravidarum, may also have higher levels of hCG, and can develop transient gestational thyrotoxicosis. As noted earlier, placental type 3 deiodinase accelerates T_4 clearance during pregnancy, which partly accounts for the increased T_4 dose requirement in treated hypothyroid women, and occasionally causes mild hypothyroidism in women with autoimmune thyroiditis and decreased thyroid gland reserve.

Transplacental delivery of thyroid-related substances can have consequences for the fetus. Maternal iodide crossing the placenta is essential for fetal thyroid hormone production; however, large amounts of iodide can actually inhibit fetal thyroid function and cause a goiter substantial enough to obstruct delivery. In women with autoimmune thyroid disease, maternal immunoglobulins that stimulate or block the TSH receptor can cross the placenta and cause thyroid dysfunction in the fetus. The antithyroid drugs methimazole and PTU cross the placenta, and in large doses will impair fetal thyroid function (see Chapter 16).

Changes in Thyroid Function with Aging

T_4 turnover is highest in infants and children, and gradually falls to adult levels after puberty. The T_4 turnover rate is then stable until after age 60, when it again begins to decline. In the elderly, the metabolic clearance of T_4 can decrease by as much as 50%, necessitating a reduction in T_4 dose for treated hypothyroid patients.

Effects of Acute and Chronic Illness on Thyroid Function (Euthyroid Sick Syndrome)

Acute or chronic illness has several striking effects on thyroid function and hormone economy. The most common and earliest effect is inhibition of T_4-to-T_3 conversion, with a resulting decrease in the circulating T_3 level. This change is accompanied by a rise in the serum rT_3 level, due to both its decreased conversion to $3,3'$-T_2 and some acceleration of its production from T_4. These changes occur physiologically in the fetus and pathologically in circumstances of caloric restriction, as in malnutrition, starvation, and anorexia nervosa, and in patients with a variety of acute or chronic systemic illnesses (Table 7–3). The pathogenesis of this *low T_3 syndrome* is thought to involve cytokines, such as tumor necrosis factor, secreted by inflammatory cells, which inhibit type 1 $5'$-deiodinase. Drugs can also inhibit type 1 $5'$-deiodinase and lower the serum T_3 levels in treated patients; these agents include corticosteroids, amiodarone, iodinated cholecystographic dyes,

TABLE 7–3 Conditions or factors associated with decreased conversion of T_4 or T_3.

1. Fetal life
2. Caloric restriction
3. Hepatic disease
4. Major systemic illness
5. Drugs:
    Propylthiouracil
    Glucocorticoids
    Propranolol (mild effect)
    Iodinated x-ray contrast agents (iopanoic acid, ipodate sodium)
    Amiodarone
6. Selenium deficiency

PTU, and high-dose propranolol. Initially, low serum T_3 levels are typically accompanied by total and free T_4 levels that are normal or, less commonly, slightly elevated (Figure 7–23). As illness becomes more severe, however, there is a decline in serum total and free T_4 as well. This hypothyroxinemia can have several causes. In certain illnesses (eg, nephrosis and severe liver disease), there can be a decline in the circulating TBG concentration. A rise in plasma free fatty acids can interfere with T_4 binding to TBG, liberating more free T_4 that, in turn, feeds back on the hypothalamus and pituitary to reduce TSH and establish a new equilibrium with a lower total T_4. TSH can also be suppressed by cytokines circulating in severely ill patients and by the dopamine or corticosteroids often used to treat critically ill patients. Finally, T_4 clearance is accelerated in illness. All of these factors can contribute to the lower serum levels of T_4 and free T_4, as measured by free-hormone immunoassays and estimated T_3 resin-uptake procedures. However, the free T_4 level measured by equilibrium dialysis has been reported to be normal in most, but not all patients.

Two clinical questions arise in sick patients with this constellation of thyroid function test abnormalities. First, it can be challenging to distinguish the euthyroid sick syndrome from central hypothyroidism; in both conditions, the T_4 and TSH levels can be low. Clinical information can be helpful, including a history of known pituitary or thyroid disease, head trauma or surgery, or
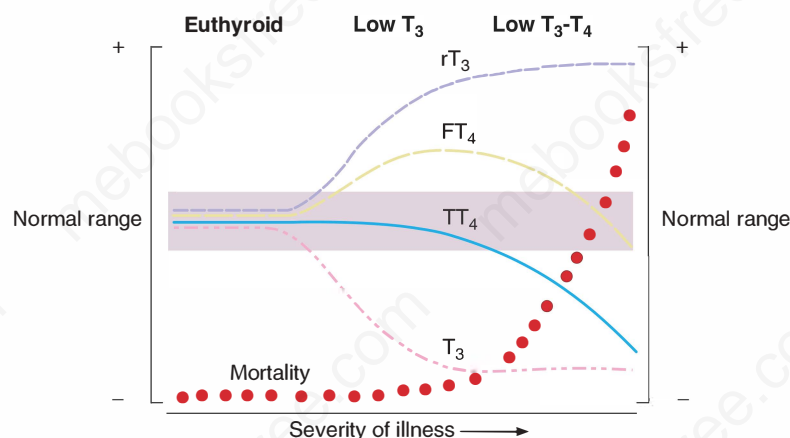

FIGURE 7–23 A schematic representation of the changes in serum thyroid hormone values with increasing severity of nonthyroidal illness. A rapidly rising mortality rate accompanies the fall in serum total T_4 (TT_4) and free T_4 (FT_4) values. (Reproduced with permission from Hurst JW. *Medicine for the Practicing Physician,* 2nd ed. Philadelphia: Butterworth-Heinemann; 1991.)

manifestations of other elements of hypopituitarism (eg, small testes in a man). Additional laboratory testing can sometimes be informative. The serum-free T_4 level determined by equilibrium dialysis is typically normal in the euthyroid sick syndrome. In postmenopausal women, an elevated serum FSH attests to at least one intact pituitary function, although gonadotropin levels can be low in severe illness as well. Measurement of serum cortisol may be particularly helpful, because it should be normal or elevated in any patient with significant systemic illness. If it is normal or elevated, hypothalamic or pituitary disease is most unlikely. Sometimes, however, it is necessary to obtain emergent imaging of the hypothalamus or pituitary to resolve the question. These thyroid function test abnormalities normalize when the patient recovers. However, recovery can be accompanied by transient rebound elevation of the serum TSH that could be misinterpreted as hypothyroidism. In this setting, in the absence of clinically apparent hypothyroidism, it is best to avoid thyroid hormone therapy and to reevaluate a few weeks following recovery.

The second clinical question is whether the euthyroid sick syndrome itself should be treated with thyroid hormone. Certainly, many critically ill patients have clinical problems that are reminiscent of hypothyroidism, such as hypothermia, altered sensorium, ventilatory and myocardial dysfunction, and hyponatremia. It has been postulated that intracellular hypothyroidism may exist in these patients and that thyroid hormone therapy might be beneficial. Conversely, it has been proposed that these changes in thyroid status are, in fact, compensatory mechanisms for reducing rates of oxygen consumption and protein catabolism and that thyroid hormone therapy might actually be harmful. Furthermore, two small randomized trials, small clinical series, and animal models of systemic illness all have failed to demonstrate any benefit of thyroid hormone therapy in this setting. Although the issue remains unresolved, current practice is generally to withhold thyroid hormone therapy, unless there is convincing clinical and biochemical evidence of true hypothyroidism.

THYROID AUTOIMMUNITY

Autoimmunity is involved in the pathogenesis of many thyroid diseases, including hyperthyroid Graves disease, hypothyroidism with autoimmune or Hashimoto thyroiditis, silent and postpartum thyroiditis, and certain forms of neonatal thyroid dysfunction. Immunologic defense against foreign organisms and substances and neoplastic cells involves macrophages that ingest and digest the foreign material and present peptide fragments on the cell surface in association with a class II protein encoded by the HLA-DR region of the *MHC* gene complex. This complex is recognized by a T-cell receptor on a CD4 helper T cell, which stimulates the release of cytokines such as interleukin-2 (IL-2). These cytokines amplify the response by inducing T-cell activation and proliferation, induction of killer cell activity in CD8 suppressor cells, and stimulation of Ab formation against the foreign antigen by B lymphocytes. Eventually, the activation process is muted by the action of the CD8 suppressor cells (see Chapter 2).

Early histopathological and experimental observations suggested the possibility that immunologic reactions could affect the thyroid gland. Thyroiditis was induced by immunization of mice with thyroid tissue and Freund adjuvant. TG Abs were identified in the sera from patients with thyroiditis, and a long-acting thyroid stimulator, later shown to be an Ab to the TSH receptor, was found in the sera of patients with Graves disease.

There are three major thyroidal antigens that can incite autoantibodies: TG, TPO, and TSH receptor. Circulating autoantibodies to these antigens are useful markers for thyroid autoimmunity, but T-cell-mediated immune mechanisms are central to the pathogenesis of thyroid diseases. Thyroid cells have the capacity to ingest antigen (eg, TG) and, when stimulated by cytokines such as gamma interferon, express cell surface class II molecules (eg, HLA-DR4) that present these antigens to T lymphocytes. In addition to the antigen and the class II molecule, other factors, such as costimulatory signals from antigen-presenting cells, are doubtless required to initiate autoimmune thyroid disease. One hypothesis is that the process is initiated or promoted by external antigens (eg, infectious agents) that lead to thyroid-directed Ab and cellular immune responses through cross-reactivity. In addition, genetic factors clearly predispose individuals to thyroid autoimmunity. HLA typing in patients with Graves disease reveals a high frequency of HLA-B8 and HLA-DR3 in Caucasians, HLA-Bw46 and HLA-B5 in Chinese, and HLA-B17 in blacks. In Caucasians, atrophic thyroiditis has been associated with HLA-B8 and goitrous Hashimoto thyroiditis with HLA-DR5. It has been suggested that a genetically-induced, antigen-specific defect in suppressor T lymphocytes may be the basis for autoimmune thyroid disease.

The increased incidence of autoimmune thyroid disease in postpubertal and premenopausal women, as well as the occurrence of postpartum thyroiditis, implies a possible role for female sex hormones or X-linked genes in the pathogenesis of autoimmune thyroid disease. A high-iodine intake may result in more highly iodinated TG that is more immunogenic and would favor the development of autoimmune thyroid disease. Therapeutic doses of lithium, used for the treatment of bipolar disorder, can interfere with suppressor cell function and may precipitate autoimmune thyroid disease.

TESTS OF THYROID FUNCTION

The high prevalence of thyroid dysfunction and neoplasia make testing for thyroid disease a daily necessity in virtually every field of medical practice—whether a pediatrician screening a neonate for congenital hypothyroidism, an internist considering possible thyroid dysfunction in a patient with nonspecific complaints, an obstetrician-gynecologist evaluating a woman with infertility, a surgeon or radiologist starting assessment of a patient with a thyroid nodule, or virtually any type of specialist considering thyroid disease as the potential cause for a complaint or test abnormality in their field of interest (eg, cardiologist, gastroenterologist, neurologist, psychiatrist, and ophthalmologist).

Several levels of laboratory and radiologic investigation for thyroid diseases should be considered. First, it is common for clinicians with a relatively low index of suspicion for thyroid dysfunction in a particular patient to want to exclude its presence

with greater certainty. For this purpose, a test with high sensitivity (ie, few false-negative results), wide availability, and low cost is most appropriate. The serum TSH concentration often plays this role in clinical consideration of hypothyroidism and hyperthyroidism. Second, when suspicion of thyroid disease is moderate to high, based on clinical or other routine laboratory results, testing with even greater sensitivity to detect even rare thyroid disorders is required, and which also begins to define the severity and underlying cause of thyroid dysfunction. For example, in a patient with hypothermia, confusion, periorbital edema, and hypercholesterolemia, measurement of both the serum TSH and free T_4 concentrations would permit exclusion of both primary and central forms of hypothyroidism; furthermore, if hypothyroidism is confirmed, the site of responsible disease would be implicated. Third, when thyroid dysfunction is clinically obvious or already biochemically confirmed by first-line testing, then more sophisticated studies are sometimes indicated to identify the underlying cause and plan appropriate therapy. For example, in a postpartum woman with thyrotoxicosis, an additional serum T_3, thyroid-stimulating immunoglobulin (TSI), and radionuclide thyroid uptake may be needed to differentiate Graves disease and postpartum thyroiditis—for which there are distinct therapeutic approaches.

TESTS OF THYROID HORMONES IN BLOOD

Serum TSH Measurement

Because thyroid dysfunction usually arises from primary disorders of the thyroid gland, serum TSH measurement is the most widely employed test to determine whether thyroid dysfunction exists. The sensitivity of the hypothalamic-pituitary-thyroid axis ensures that both primary hypothyroidism and thyrotoxicosis, due to either primary thyroid disorders or exogenous thyroid hormones, are detected. Furthermore, even mild degrees of thyroid dysfunction, in which the serum thyroid hormone concentrations have fallen or risen significantly but remain within the broad reference range (ie, subclinical or mild thyroid dysfunction) are identified. The inverse logarithmic relationship between serum TSH and free T_4 concentration in various situations is shown in Figures 7–24A and 7–24B.
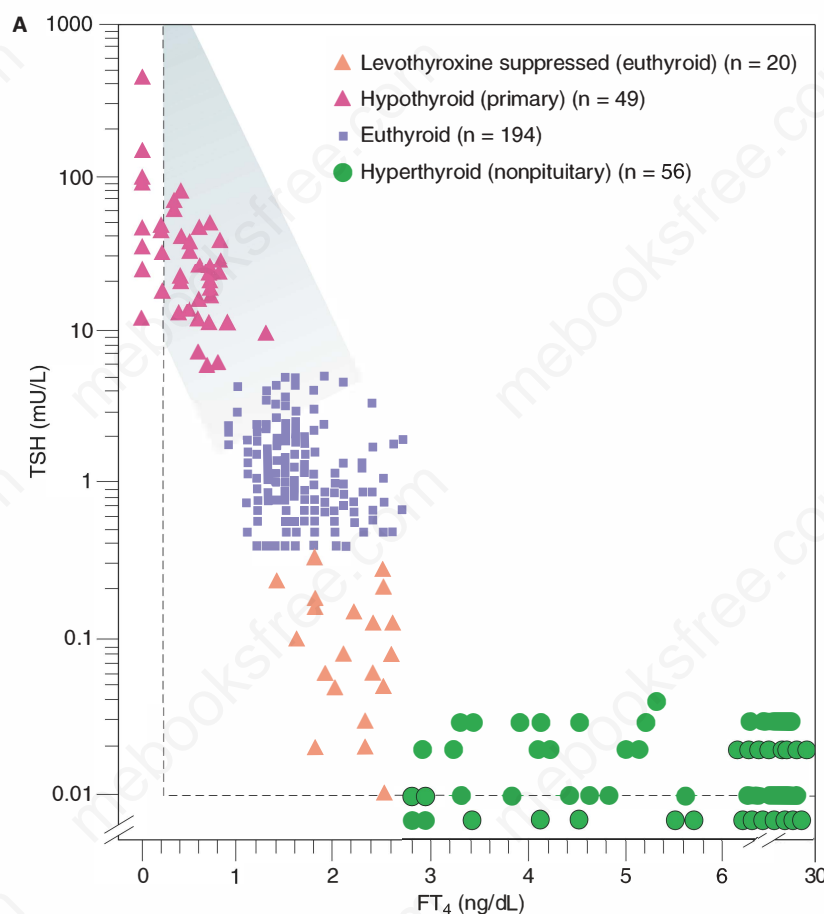
FIGURE 7–24 **A.** Relationship between serum-free thyroxine by dialysis (FT_4) ng/dL and $\log_{10}$ TSH in euthyroid, hyperthyroid, hypothyroid, and T_4-suppressed euthyroid individuals. Note that for each unit change in T_4 there is a logarithmic change in TSH. **B.** Relationship between FT_4 by dialysis and $\log_{10}$ TSH in normal subjects and subjects with various illnesses including primary hypothyroidism, central hypothyroidism and non–levothyroxine-suppressed TSH, levothyroxine-treated congenital hypothyroidism, TSH-secreting tumors, thyroid hormone resistance, nonthyroidal illness, and nonpituitary hyperthyroidism. (Reproduced with permission from Kaptein EM. Clinical applications of free thyroxine determinations. *Clin Lab Med.* 1993 Sep;13(3):653-672.)
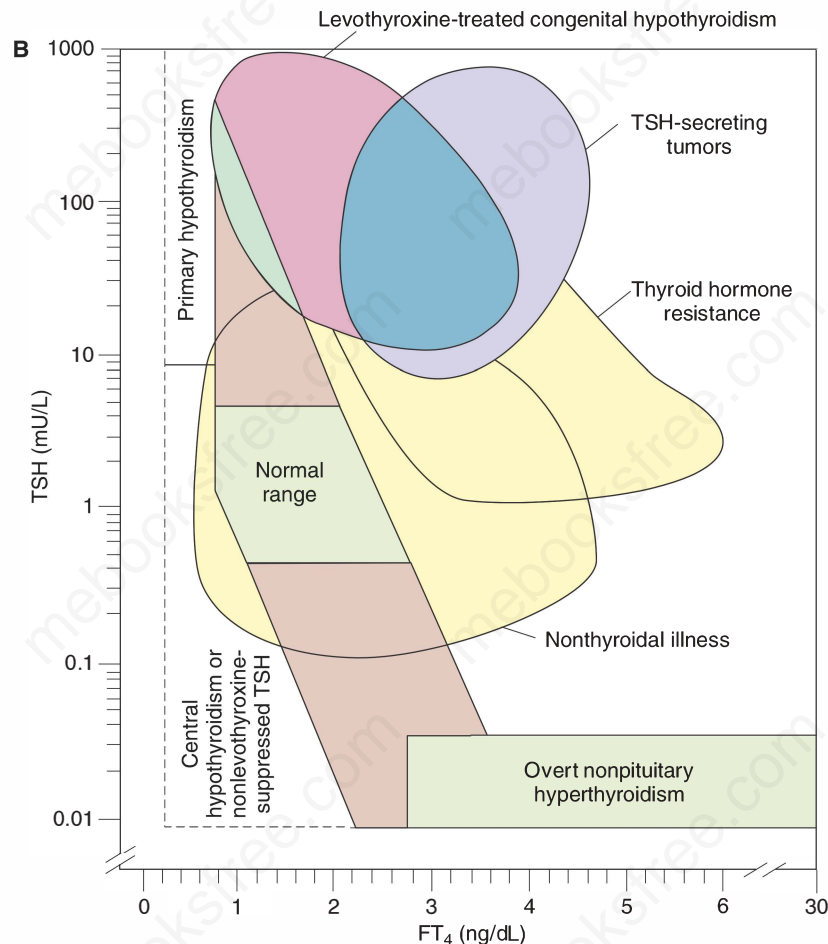
FIGURE 7–24 *(Continued)*

TSH assays must be sensitive (ie, able to distinguish low from very low and undetectable TSH levels) to be used for diagnosis of thyrotoxicosis. A National Academy of Clinical Biochemistry guideline specifies that sensitivity, or lower limit of detection, for TSH assays should be less than 0.02 mU/L. It is better to assess and compare TSH assays by this quantitative criterion rather than marketing terms such as *third-generation* or *ultrasensitive*.

There are, in general, two sorts of immunoassays used to measure TSH (and T_4 and T_3) in serum samples: immunometric assays (IMA) and radioimmunoassay (RIA). In an IMA or *sandwich assay*, an Ab—usually a mouse monoclonal Ab—directed against one epitope on the TSH molecule is fixed to a solid matrix, and a second monoclonal TSH Ab that binds another TSH epitope is labeled with a detectable marker, which can be a radioisotope, a colorimetrically quantifiable enzyme, or a fluorescent or chemiluminescent tag. In such IMAs, the concentration of TSH, which links the solid state and labeled Abs, is proportionate to the intensity of signal emitted by the marker once unbound second Ab has been separated off. In contrast, in conventional RIAs, a small amount of a TSH tracer, to which a radioactive molecule has been linked, competes for binding to first Ab (eg, a rabbit antihuman TSH polyclonal Ab). Then Ab-bound TSH, both in the sample and the tracer, is separated from the free tracer in the supernatant using one of several techniques: a second Ab directed against the first (eg, goat antirabbit immunoglobulin Ab), polyethylene glycol, or staphylococcal protein A. In RIAs, the concentration of TSH in the sample is inversely proportionate to tracer activity. In general, TSH IMAs are more sensitive and widely employed than RIAs.

There are infrequent but important limitations of the serum TSH assay alone for diagnosis of thyroid dysfunction when clinical suspicion of hypothyroidism is high and in certain other special circumstances (Table 7–4). First, central hypothyroidism can be overlooked. In patients with hypothyroidism due to a pituitary tumor, for example, the serum TSH is typically not markedly elevated, but rather low or inappropriately normal or minimally elevated. Consequently, simultaneous measurements of both the serum-free T_4 and TSH would be required to further substantiate or exclude this diagnosis. Central hypothyroidism can also be confirmed by assessing the serum TSH response to an injection of TRH. To perform this test, serum samples for TSH are obtained before and 30 to 60 minutes after 200 µg of TRH is administered intravenously. The absence of a significant rise in serum TSH concentration (eg, to a peak value >7 mU/L) indicates pituitary or hypothalamic disease or suppression by exogenous thyroid hormone. A modest and delayed rise may be seen in patients with

TABLE 7–4   Causes of inappropriate serum TSH concentration.[a]

Clinical Hyperthyroidism
TSH-secreting pituitary adenoma
Isolated pituitary resistance to thyroid hormone
Clinical Hypothyroidism
Central (pituitary or hypothalamic) hypothyroidism
Preceding TSH suppression (eg, recently treated hyperthyroidism)
Clinical Euthyroidism
Systemic illnesses ($\downarrow$ during acute phase, $\uparrow$ during recovery)
Generalized resistance to thyroid hormone (compensated)
Assay interference
Anti-TSH antibodies
Anti-mouse immunoglobulin antibodies
Drugs: dopamine, dobutamine, glucocorticoids
Normal effect of aging, especially age >80 years
Marked obesity

[a]"Inappropriate" refers to disruption of the usual reciprocal relationship between serum TSH and free thyroid hormone concentrations.

hypothalamic disease and tertiary hypothyroidism. Currently, TRH is not available in the United States. Another cause of a misleadingly low serum TSH concentration is drug treatment with either high-dose corticosteroids or dopamine, both of which inhibit TSH secretion (see Table 7–4).

A second limitation of TSH testing alone is that there are other nonthyroidal causes of an elevated serum TSH concentration, as described in Table 7–4.

Serum T_4 and T_3 Measurements

The concentrations of total and free (unbound) T_4 and T_3 are measurable by a variety of automated assay techniques. Serum total thyroid hormone concentrations are widely available and accurate for assessment of most patients with overt thyroid dysfunction. However, a limitation of these assays is they can provide false-positive and false-negative results when changes in the concentration or binding affinity of the thyroid hormone–binding plasma proteins are altered. For example, a pregnant woman with an estrogen-induced increase in the concentration of TBG could be misdiagnosed as being hyperthyroid if only a total T_4 assay was employed. Conversely, a pregnant patient with primary hypothyroidism might be overlooked if the elevated TBG concentration produced total T_4 within the reference range. Despite these reservations, diagnostic accuracy is high when total T_4 assay results are interpreted along with TSH assay findings.

These problems are circumvented by measuring the free thyroxine (FT_4) concentration by free T_4 IMAs, equilibrium dialysis, or liquid chromatography/mass spectrometry (LC/MS). Free T_4 assays employ strategies to measure only the unbound fraction of T_4 in serum samples. One approach is to employ a competing T_4 analog tracer that binds to the assay Ab but not to plasma proteins. Such free T_4 assays are quite accurate in distinguishing straightforward TBG derangements from true FT_4 abnormalities due to thyroid dysfunction. However, they may be influenced to some extent by variations in total T_4, and they can yield misleadingly high readings in dysalbuminemic hyperthyroxinemia and

falsely low readings in the third trimester of pregnancy. However, free T_4 IMAs are currently the most widely used method for free T_4 determination because they are accurate in most settings, rapid, automatable, and inexpensive.

Equilibrium dialysis entails equilibrating a serum sample with buffer across a membrane with pores that permit passage only of unbound T_4; then the dialysate is assayed for its T_4 concentration, from which the serum concentration can be derived. Although equilibrium dialysis is considered the gold standard for free T_4 determination, it is cumbersome, relatively expensive, and not widely available. Consequently, it is not the optimal first-line method for free T_4 measurement, and it is best reserved for special circumstances.

Free T_4 estimation by free T_4 index entails measuring the total T_4 in a conventional IMA and then estimating the partition of a tracer iodothyronine between plasma proteins and a matrix with a virtually limitless binding capacity for small molecules (resin, charcoal, or talc). Radiolabeled T_3 is often the tracer employed, although T_4 tracer can be used as well. The free T_4 index is then calculated as the product of the total T_4 multiplied by the percentage of iodothyronine tracer taken up by the matrix, the so-called T_3 uptake. It is important to note that this refers to the T_3 tracer used in the assay, not the T_3 level in the serum sample.

When the free T_4 concentration alone is used to diagnose thyroid dysfunction, there are a number of conditions from which true hyperthyroidism or hypothyroidism must be distinguished. First, there are several conditions in which euthyroid individuals can develop hyperthyroxinemia (see Table 7–1). These include rare patients with nonthyroidal illnesses; unusual plasma-binding proteins abnormalities, such as familial dysalbuminemic hyperthyroxinemia (see later); and illness or drugs that block T_4-to-T_3 conversion (eg, iodinated contrast media, amiodarone, glucocorticoids, and propranolol). In these conditions, true hyperthyroidism is excluded by the normal serum TSH level. Conversely, there are also conditions in which the serum-free T_4 can be subnormal in euthyroid individuals. Antiepileptic drugs, such as phenytoin and carbamazepine, have multiple actions that lower total and free T_4 levels, including inhibition of TSH secretion and accelerating T_4 catabolism. However, the normal accompanying serum TSH and equilibrium dialysis T_4 levels implies that these patients are, in fact, euthyroid.

Total and free T_3 concentrations can also be quantitated in specific IMAs. Measurement of serum T_3 is used (1) to recognize patients who have T_3 toxicosis, a milder degree of hyperthyroidism in which the serum T_3 level is elevated with a normal serum T_4; (2) to fully define the severity of hyperthyroidism and monitor the response to therapy; and (3) to help in the differential diagnosis of patients with hyperthyroidism. Because T_3 is preferentially secreted in most patients with Graves disease, and some with toxic nodular goiter, the ratios of serum T_3:T_4 (expressed in ng/dL:µg/dL) are typically greater than 20 in patients with these conditions. In contrast, in thyroiditis-related thyrotoxicosis the serum T_3:T_4 ratio is usually less than 20, reflecting the relative amounts of the two hormones stored in and leaking from the inflamed gland.

The serum T_3 concentration is not accurate for diagnosis of hypothyroidism because a normal T_3 level is often maintained in

TABLE 7–5 Representative iodothyronine kinetic values in a euthyroid human.

	T_4	T_3	rT_3
Serum levels			
Total, µg/dL (nmol/L)	8 (103)	0.12 (1.84)	0.04 (0.51)
Free, ng/dL (pmol/L)	1.5 (19)	0.28 (4.3)	0.24 (3.69)
Body pool, µg (nmol)	800 (1023)	46 (70.7)	40 (61.5)
Distribution volume (L)	10	38	98
Metabolic clearance rate (MCR) (L/d)	1	22	90
Production (disposal) rate MCR × serum concentration, µg/d (nmol/d)	80 (103)	26 (34)	36 (46)
Half-life in plasma ($t1/2$) (days)	7	1	0.2

Note: T_4 µg/dL × 12.87 = nmol/L; T_3 µg/dL × 15.38 = nmol/L.

FIGURE 7–25 Typical curves of 24-hour radioiodine uptake in normal subjects and in patients with thyroid disease. A high-turnover curve may be seen in patients taking an iodine-deficient diet or with a defect in hormone synthesis.

patients with mild to moderate primary hypothyroidism, in part because TSH stimulation of the thyroid gland increases the relative secretion of T_3. Because T_3 is bound to TBG, the total serum T_3 concentration varies with the TBG level, although less than the serum total T_4 (Table 7–5). Serum-free T_3 can be measured by IMA or more precisely by equilibrium dialysis.
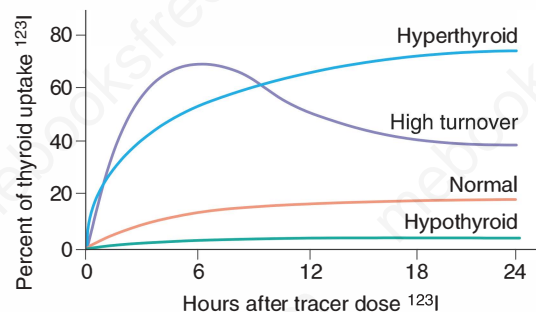
Reverse T_3 can be measured by RIA; its serum concentration in adults is about one-third of the total T_3 concentration. However, there is little or no clinical indication for rT_3 measurement. Previous claims that it is useful to differentiate hypothyroidism from the thyroid function test changes of nonthyroidal illness have not shown that rT_3 levels are sufficiently accurate for that purpose.

TG can be measured in serum by either IMA or RIA, which now have sensitivities as low as 0.1 ng/mL. There are two indications for TG measurement: (1) detection of residual or recurrent epithelial thyroid cancers (papillary, follicular, and Hürthle cell carcinomas) after thyroidectomy (see section Management of Thyroid Cancer); and (2) differentiation of thyrotoxicosis due to exogenous thyroid hormone, in which the TG level is suppressed, from all forms of endogenous hyperthyroidism, including various forms of thyroiditis, in which it is elevated or normal.

There are several potential limitations of current TG IMAs. The most important is interference caused by the presence of circulating TG Abs. This usually leads to spuriously low values in thyroglobulin IMAs and falsely elevated values in RIAs. Serum TG levels are not useful to differentiate malignant from benign thyroid nodules.

ASSESSMENT OF THYROID IODINE METABOLISM AND BIOSYNTHETIC ACTIVITY

Radionuclide uptake measurements can be used to differentiate among the causes of thyrotoxicosis, particularly when used in conjunction with imaging. Radioiodine uptake determinations

permit *in vivo* assessment of fractional iodine uptake and turnover by the thyroid gland. Iodine-123 is the ideal isotope for this purpose, with 13-hour half-life and emitting quantifiable 28-keV x-rays and 159-keV gamma photons but no tissue damaging beta particles. [123]I—or when it is unavailable, a low dose of [131]I—is given orally, and radioactivity over the thyroid is quantified with a scintillation counter at 4 or 6 hours and again at 24 hours (Figure 7–25). The fractional radioactive iodine uptake (RAIU) by the thyroid varies with recent dietary iodide intake. In countries with a relatively high iodide intake (eg, USA, Chile, and Brazil), the normal RAIU is 5% to 15% at 6 hours and 8% to 30% at 24 hours, whereas in areas of low-iodide intake and endemic goiter (eg, some parts of Europe and sub-Saharan Africa), the 24-hour RAIU can be as high as 60% to 90%.

In hyperthyroid Graves disease, the 24-hour RAIU is markedly elevated (eg, 40%-95%), although if iodide turnover is very rapid, the 5-hour uptake may be even higher than the 24-hour uptake (see Figure 7–25). Similarly, the 24-hour RAIU is increased to a lesser extent in toxic nodular goiter (eg, 30%-70%). Thyrotoxicosis with a low-thyroidal RAIU occurs in patients with several conditions: (1) subacute and lymphocytic (postpartum, silent) thyroiditis with inflammatory release of preformed hormone causing spontaneously resolving thyrotoxicosis; (2) in iatrogenic or factitious thyrotoxicosis resulting from excessive thyroid hormone medication; and (3) following excess oral or intravenous iodide intake, as occurs, for example, with amiodarone therapy or after radiocontrast agents. There can also be low RAIU over the anterior neck when there is excessive thyroid hormone production elsewhere in the body, as occurs with ectopic thyroid tissue at the tongue base (lingual thyroid); in the mediastinum; with struma ovarii, an ovarian teratoma with functioning thyroid tissue; and with functioning metastases from thyroid carcinoma after total thyroidectomy.

THYROID IMAGING

1. RADIONUCLIDE IMAGING

[123]I and technetium ^{99m}Tc pertechnetate ([99m]TcO$_4$) are useful for determining the functional activity and morphology of the
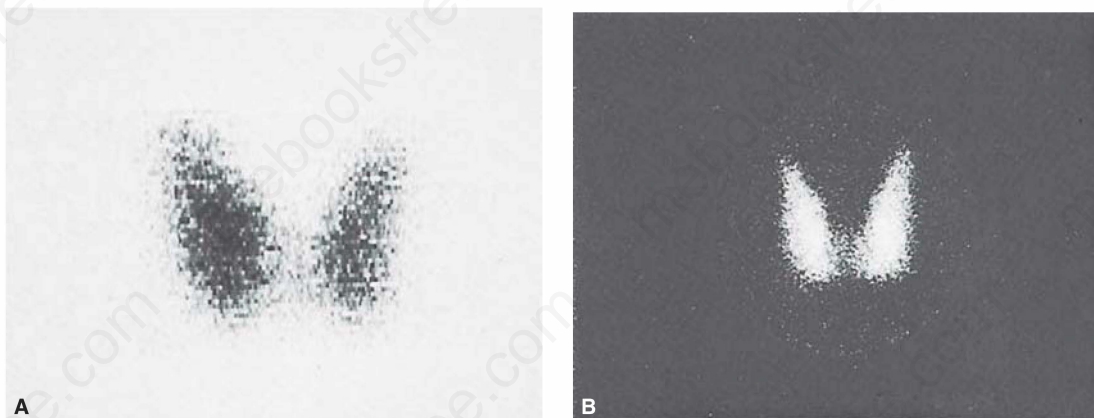
FIGURE 7–26 **A.** Rectilinear sodium ^{123}I scan performed 6 hours after the ingestion of 100 μCi of sodium [123]I. **B.** Scintiphoto (pinhole collimator) thyroid scan performed 6 hours after the ingestion of 100 μCi of sodium [123]I. (Used with permission from Dr. RR Cavalieri.)

thyroid gland. [123]I (200-300 μCi) is administered orally, and an image of the thyroid is obtained 8 to 24 hours later. $^{99m}TcO_4$ (1-10 mCi), which is trapped by the NIS, but not organified and retained in the gland, is administered intravenously, and the scan image is obtained earlier than usual—at 30-60 minutes. Images can be obtained with either a rectilinear scanner or a gamma camera. The rectilinear scanner moves back and forth over the area of interest and produces a life-size picture, on which regions of interest, such as nodules, can be marked (Figure 7–26A). The gamma camera has a pinhole collimator, and the scan is obtained on a fluorescent screen and recorded on film or a computer monitor. The camera has greater resolution, but regions of interest must be identified with a radioactive marker for clinical correlation (Figure 7–26B).

Radionuclide scans provide information about the size and shape of the thyroid gland and the distribution of tracer activity within the gland. This is useful in differentiating among the causes of thyrotoxicosis. For example, in patients with hyperthyroidism, an enlarged gland with intense and homogenous concentration of tracer is typical of Graves disease. In contrast, one or more discrete regions of tracer activity, especially corresponding to a palpable nodule or nodules, with suppression of extranodular tissue are characteristic of toxic nodular goiter.

Radionuclide imaging has a limited role in differentiation of benign from malignant thyroid nodules. When a patient with a thyroid nodule has a suppressed serum TSH concentration, a radionuclide scan is often appropriate to determine if the nodule is *hot* (ie, functioning with suppression of extranodular thyroid tissue). If this is the case, the nodule can generally be considered benign. Because both most benign and malignant thyroid nodules are *cold*, or hypofunctioning, documenting this is not particularly helpful, and cytologic evaluation must be performed in any case.

Radionuclide whole body scanning is useful in follow-up of patients with treated thyroid cancer and for confirmation that a mass on the tongue (lingual thyroid), in the midline of the neck (thyroglossal duct), mediastinum (substernal goiter), or pelvis (struma ovarii) represents ectopic thyroid tissue.

THYROID ULTRASONOGRAPHY AND OTHER IMAGING TECHNIQUES

The size of the thyroid gland and characteristics of nodular lesions within it are best ascertained by high-resolution ultrasonography (Figure 7–27). Sonography can differentiate solid nodules from cystic lesions and characterize complex cysts that have both solid and cystic components. Certain features of nodules are more common in malignant lesions (eg, hypoechogenicity, irregular nodule capsule, and microcalcifications), and other features are typical of benign lesions (eg, iso- or hyperechogenicity, a *spongiform* appearance and cystic change). Sonography can be used to serially monitor the size of thyroid nodules over time, to guide fine-needle aspiration of thyroid nodules, and to assess regional lymph nodes in thyroid cancer patients both preoperatively and in long-term follow-up. Sonographic assessment of thyroid gland blood flow can be useful in the differential diagnosis of amiodarone-induced thyrotoxicosis.

Computed tomography (CT) and magnetic resonance imaging (MRI) are useful in patients with thyroid disease in special settings. They may be required to define the posterior cervical and substernal extent of a large goiter, and whether there is tracheal deviation or compression. In thyroid cancer patients, they can be helpful in identifying metastatic disease outside of the neck (eg, in the lungs, liver). However, it should be remembered that administration of iodine-containing radiocontrast dye for CT imaging interferes with subsequent radioiodine imaging and therapy for up to several months. Positron emission tomography (PET) scanning has a role in localization of metastatic disease in treated thyroid cancer patients who have a serum TG concentration greater than 10 ng/mL.

When any of the earlier-described techniques are used for other indications to image structures in the cervical region, thyroid nodules are commonly detected incidentally. Studies indicate that the prevalence of thyroid cancer in such thyroid incidentalomas is the same as in palpable nodules, so they should usually be biopsied if greater than 1.0 to 1.5 cm in diameter, and monitored by serial ultrasound if smaller. FDG PET incidentalomas, in particular, should be thoroughly evaluated, as some studies have shown that many of these hypermetabolic lesions have been found to be cancers.

Glands afflicted with thyroiditis can also be identified on FDG PET scanning.

THYROID BIOPSY

Fine-needle aspiration biopsy (FNAB) of a thyroid nodule to obtain material for cytopathological inspection is the best method for differentiating benign from malignant thyroid nodules and diffuse goiters. FNAB is a minor outpatient procedure performed in awake, nonfasting patients with appropriate aseptic technique; local anesthesia may or may not be required. The procedure is typically performed with a 27- or 25-gauge needle, which is inserted into the nodule or goiter, and moved in and out until a small amount of bloody material is seen in the hub of the needle. The aspirated material is then expressed onto a glass slide and prepared as a thin smear, which is either air-dried for immediate
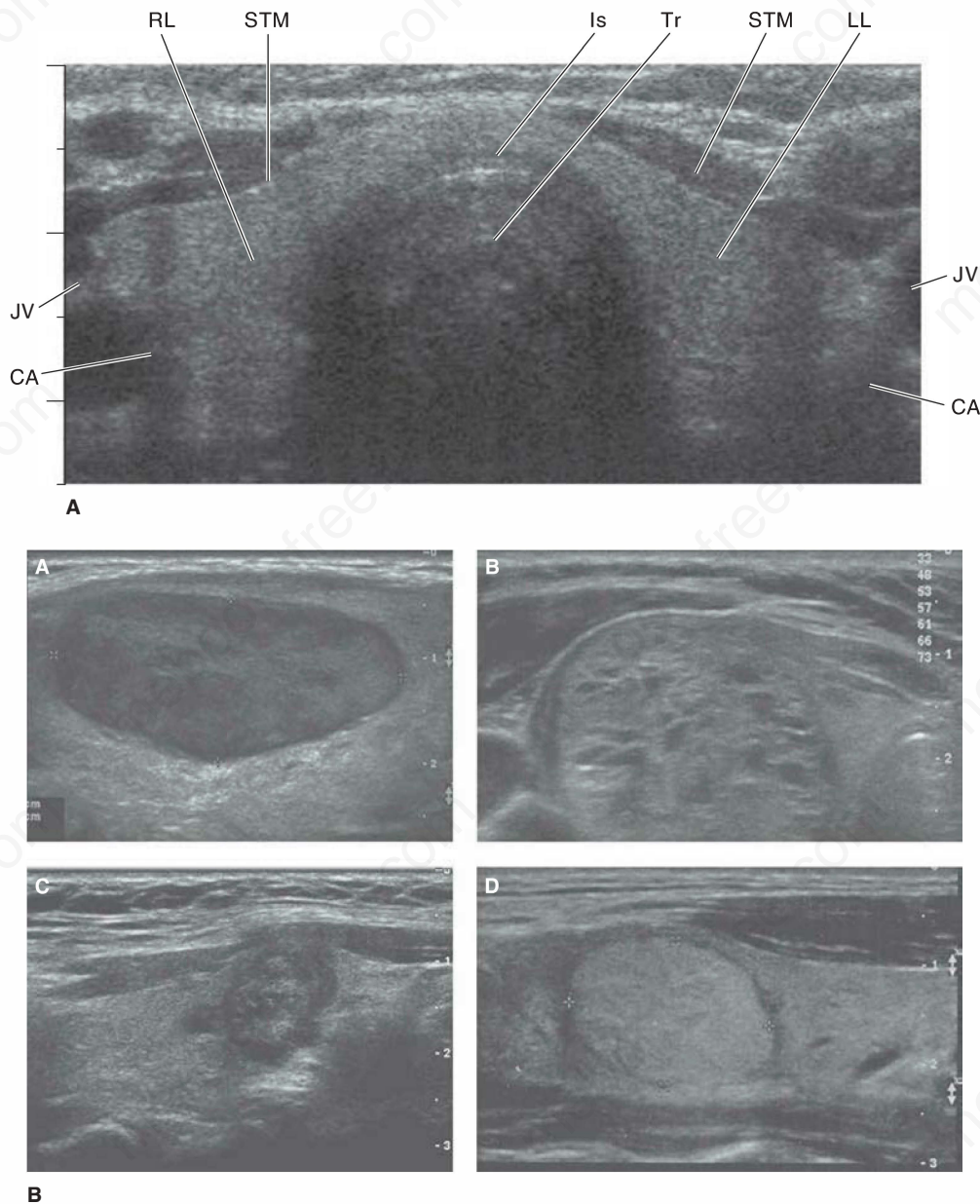
FIGURE 7–27 Sonographic patterns and risk of malignancy. **A.** Ultrasound of the normal thyroid gland. (Tr, trachea; LL, left lobe of thyroid; RL, right lobe of thyroid; Is, isthmus; CA, carotid artery; JV, jugular vein; STM, sternothyroid muscle.) (Used with permission from Dr. RA Filly.) **B.** Ultrasound of the thyroid illustrating various sonographic features of thyroid nodules. A: Hypoechoic nodule; B: Spongiform nodule; C: Thyroid cancer showing irregular margins and microcalcifications; D: Hyperechoic nodule. **C.** Ultrasound of the thyroid illustrating a large thyroid cyst (A) and Graves' disease (B) with enlarged heterogeneous echotexture and markedly increased vascular flow on Doppler ultrasound (C). (Reproduced with permission from Baskin HJ, Duick DS, Levine RA. *Thyroid Ultrasound and Ultrasound-guided FNA*, 3rd ed. New York: Springer Verlag; 2013.)

Trans is thmus
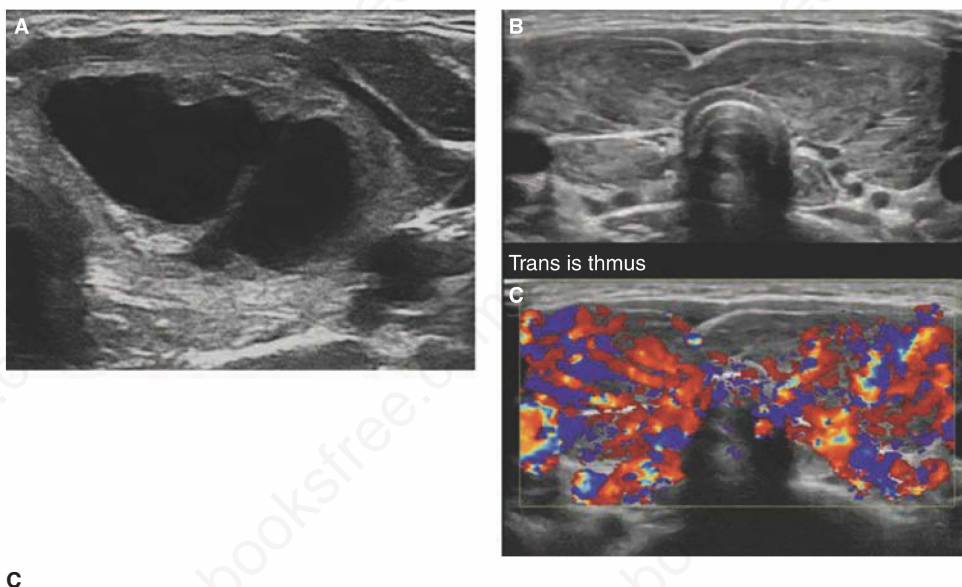
FIGURE 7–27 *(Continued)*

staining and preliminarily assessed, or placed in fixative for definitive staining and cytopathologic assessment. FNAB can also be useful to evacuate cystic lesions, although the fluid recollects in one-half of these cysts.

Cytologic diagnosis of malignancy in thyroid nodules is highly accurate. The sensitivity for cancer detection (true-positive diagnoses divided by total number of nodules proven to be cancer by surgery) is 95% to 98%, and the specificity for cancer (true-positive results divided by total number of cases with a positive biopsy) is also 98%. Conversely, among the remaining nodules, a benign cytological finding is very reassuring with a 97% negative predictive value (true-negative results divided by all negative results). However, approximately one-third of thyroid nodules yield tissue that is cytologically indeterminate, which can be subcategorized as atypia of undetermined significance, follicular neoplasm, or suspicious for malignancy with estimated cancer risks of 5% to 15%, 15% to 30%, and 60% to 75%, respectively. These issues are discussed in depth in the section Thyroid Nodules and Thyroid Cancer.

TEST OF PERIPHERAL THYROID HORMONE ACTIONS

Theoretically, the most definitive test of thyroid function would be a method of quantifying thyroid hormone responses in peripheral tissues. In practice, however, these tests have limited accuracy because the response parameters may be imprecisely quantified (eg, ankle reflex relaxation time) or because they are also modulated by nonthyroidal factors (eg, serum cholesterol and echocardiographic systolic time intervals). Nonetheless, these techniques are useful in clinical research and in assessment of rare patients suspected of having resistance to thyroid hormone (RTH).

Basal oxygen consumption reflects thyroid hormones' effects on oxygen consumption and calorigenesis in most body tissues, especially the liver. There is considerable overlap between the measured levels in hypothyroidism, euthyroidism, and hyperthyroidism. Echocardiography can be used to assess systolic and diastolic functions, both of which are altered with thyroid dysfunction; but this is exclusively a research technique.

Thyroid hormones influence the concentrations of a number of analytes in blood. For example, serum total and LDL cholesterol is usually elevated in hypothyroidism and lowered in hyperthyroidism. In hyperthyroidism, sex hormone–binding globulin, ferritin, and angiotensin-converting enzyme concentrations are often higher, as are markers of bone formation and degradation. However, none of these test results is sensitive or specific enough for routine diagnostic use.

MEASUREMENT OF THYROID AUTOANTIBODIES

Detection of thyroid autoantibodies can be helpful in establishing the diagnosis of autoimmune thyroid disease. To establish the diagnosis of autoimmune (Hashimoto) thyroiditis in a patient with diffuse goiter or hypothyroidism, serum anti-TPO or the less sensitive anti-TG Abs can be employed. An older test, antimicrosomal Abs has intermediate sensitivity. Anti-TPO and anti-TG Abs can be measured by hemagglutination, enzyme-linked immunosorbent assay (ELISA), or RIA; RIA is more sensitive, precisely quantifiable, and generally preferred. Although these Abs are most typical of autoimmune thyroiditis, they are also detected in significant proportions of patients with Graves disease and certain other forms of thyroiditis. Due to their high prevalence in the population, especially in women, a positive result does not rule out presence of a second thyroid disorder (eg, cancer in an asymmetrical goiter).

TSH receptor–stimulating, –binding, and –blocking Abs can all be detected. TSI (also called TSAb) is present in approximately 90% of patients with Graves disease and is almost never detected in unaffected patients or those with other autoimmune thyroid diseases. The test is performed by exposing a patient's IgG to cultured Chinese hamster ovary cells that have been transfected with the TSH receptor gene and then measuring their cAMP response in comparison with control IgG and TSH standards. TSI testing is used in a few special settings: (1) pregnant women to define risk of subsequent neonatal hyperthyroidism in their fetus and neonate; (2) to differentiate Graves disease from postpartum thyroiditis in women who are breast feeding and cannot have a radionuclide thyroid uptake; (3) to make the diagnosis of Graves disease in euthyroid patients with apparent thyroid-related ophthalmopathy; and (4) to predict whether antithyroid drug-treated Graves patients will relapse when drug is later discontinued. TSH receptor–binding immunoglobulins can also be detected by *in vitro* demonstration that circulating immunoglobulins compete with radiolabeled TSH in binding to recombinant TSH receptors. It is a simpler alternative for the TSI indications described earlier, but does not distinguish between stimulating and blocking anti-TSH receptor Abs. TSH receptor blocking Abs have been implicated as the cause of hypothyroidism in a minority of hypothyroid patients in certain populations; they can also be transmitted transplacentally to the fetus and inhibit thyroid function during pregnancy and in the neonatal period.

DISORDERS OF THE THYROID

Patients with thyroid disease usually complain of one or more of the following: (1) thyroid enlargement, which may be diffuse or nodular; (2) symptoms of thyroid hormone deficiency, or hypothyroidism; (3) symptoms of thyroid hormone excess, or hyperthyroidism; or (4) complications of a specific form of hyperthyroidism—Graves disease—that may present with striking prominence of the eyes (exophthalmos) and, rarely, thickening of the skin over the lower legs (thyroid dermopathy or pretibial myxedema). Patients with thyroid nodules, whether benign or malignant, often have no symptoms, as do persons with mild perturbations of thyroid function.

History

The history should include evaluation of symptoms related to an enlarged gland, thyroid hormone deficiency and excess, as discussed in more detail later. The history should identify risk factors for thyroid disease. Exposure to ionizing radiation in childhood has been associated with an increased incidence of thyroid disease, including cancer. Iodide ingestion in the form of amiodarone, an iodine-containing antiarrhythmic drug, or intravenous iodide-containing contrast media used in angiography or CT scanning may induce hypothyroidism or hyperthyroidism. Residence in an area of low dietary iodide is associated with iodine deficiency goiter (endemic goiter). The family history should be explored with particular reference to goiter, hyperthyroidism, hypothyroidism, or thyroid cancer as well as immunologic disorders such as type 1

diabetes, pernicious anemia, alopecia, vitiligo, or myasthenia gravis, which may be associated with an increased incidence of autoimmune thyroid disease. Multiple endocrine neoplasia types 2A (Sipple syndrome) and 2B with medullary carcinoma of the thyroid gland are autosomal dominant conditions.

Physical Examination

Physical examination of the thyroid gland is illustrated in Figure 7–28. The thyroid is firmly attached to the anterior trachea midway between the sternal notch and the thyroid cartilage; it is often easy to see and to palpate. The patient should have a glass of water for comfortable swallowing. There are three maneuvers:

1. With a good light coming from behind the examiner, the patient is instructed to swallow a sip of water. Observe the gland as it moves up and down. Enlargement and nodularity can often be noted.
2. Palpate the gland anteriorly with swallowing. The thumb placed anteriorly along the trachea allows localization of the isthmus and connected lobes, and if present the pyramidal lobe, which extends superiorly from the isthmus and is often palpable in patients with autoimmune thyroid disease.
3. Then, the examiner's right hand is placed on the patient's left shoulder, and the examiner's right thumb is used to palpate the left lobe, as the patient swallows. This permits a full appreciation of the size, texture, and contour of the lobe. Then, using the left hand, the right lobe can be palpated. The examiner's thumbs should be medial to the sternocleidomastoid muscles to avoid mistaking them for the thyroid lobes. Some experts recommend palpation from behind using three fingers to palpate each lobe while the patient swallows, but this technique has the disadvantage of not being able to see the gland during palpation.

Normally, the thyroid is just barely palpable, has a smooth surface, and has a soft to rubbery consistency. The palpable bulbous portion of each lobe of the normal thyroid gland measures about 2 cm in vertical dimension and about 1 cm in horizontal dimension above the isthmus. An enlarged thyroid gland is called a **goiter**. Generalized enlargement is termed **diffuse goiter**; irregular or lumpy enlargement is called **nodular goiter**. In Hashimoto thyroiditis, the gland is often symmetrically enlarged, firm, and has a bosselated (cobblestone) or finely nodular surface. In the atrophic form of Hashimoto thyroiditis, the gland may not be palpable at all. In patients with Graves disease, the gland is usually, but not always, symmetrically enlarged, smooth, and rubbery in consistency. Multinodular goiters have one or more distinct nodules palpable, although both small and large and nodules may not be palpable because of their location or consistency. Tenderness of the gland should be noted. In patients with Graves disease, the presence of a bruit by auscultation should be sought, and a thrill may be palpable over the gland. Tracheal deviation by a goiter, cervical lymphadenopathy possibly related to thyroid cancer, and jugular venous distention and facial erythema from thoracic inlet obstruction by a large goiter should be noted as well. An outline of the gland can be traced on the skin of the neck and measured (Figure 7–28D). Nodules can be measured in a similar way. Thus, changes in the size of the gland or in nodules can easily be followed.
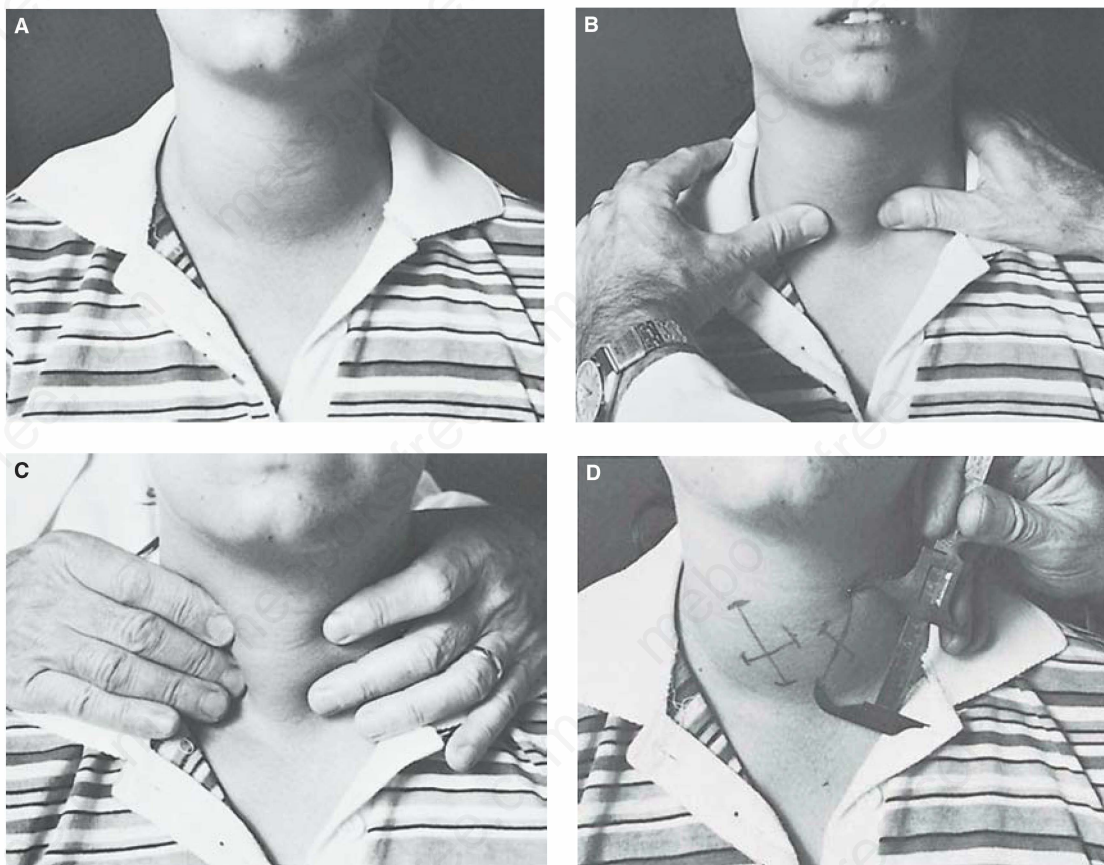
FIGURE 7–28 Examination of the thyroid gland. **A.** Observe the neck, especially as the patient swallows. **B.** Examine from the front, rotating the gland slightly with one thumb while palpating the other lobe with the other thumb. **C.** Examine from behind, using three fingers and the same technique. **D.** The size of each lobe or of thyroid nodules can be measured by first drawing an outline on the skin.

HYPOTHYROIDISM

Hypothyroidism is a clinical syndrome resulting from a deficiency of thyroid hormones, which results in widespread organ-specific effects. Hypothyroidism in infants and children results in marked slowing of growth and development, with serious permanent consequences, including mental retardation when it occurs in infancy and short stature, when it occurs in later childhood. Hypothyroidism with onset in adulthood causes diminished calorigenesis and oxygen consumption; impaired cardiac, pulmonary, renal, gastrointestinal, and neurological functions and deposition of glycosaminoglycans in intracellular spaces, particularly in skin and muscle, producing in extreme cases the clinical picture of **myxedema**. The symptoms and signs of hypothyroidism in adults (described in detail later) are reversible with therapy.

Etiology and Incidence (Table 7–6)

Hypothyroidism may be classified as (1) *primary* which is by far the most common, (2) *secondary* due to pituitary TSH deficiency, (3) *tertiary* due to hypothalamic TRH deficiency, or (4) *peripheral thyroid hormone resistance*.

The incidence of various causes of hypothyroidism varies depending on geographic and environmental factors, such as

TABLE 7–6 Etiology of hypothyroidism.

Primary:
1. Hashimoto thyroiditis:
    a. With goiter
    b. "Idiopathic" thyroid atrophy, presumably end-stage autoimmune thyroid disease, following either Hashimoto thyroiditis or Graves disease
    c. Neonatal hypothyroidism due to placental transmission of TSH-R blocking antibodies
2. Radioactive iodine therapy for Graves disease
3. Subtotal thyroidectomy for Graves disease, nodular goiter, or thyroid cancer
4. Excessive iodide intake (kelp, radiocontrast dyes)
5. Subacute thyroiditis (usually transient)
6. Iodide deficiency (rare in North America)
7. Inborn errors of thyroid hormone synthesis
8. Drugs
    a. Lithium
    b. Interferon-alfa
    c. Amiodarone

Secondary: Hypopituitarism due to pituitary adenoma, pituitary ablative therapy, or pituitary destruction

Tertiary: Hypothalamic dysfunction (rare)

Peripheral resistance to the action of thyroid hormone

dietary iodide and goitrogen intake, the genetic characteristics of the population, and the age distribution of the population (pediatric or adult). Iodine deficiency is is still frequently seen in developing countries and is the most common cause of hypothyroidism world wide. The causes of hypothyroidism, listed in approximate order of frequency in the United States, are presented in Table 7–6. Hashimoto thyroiditis is by far the most common cause of hypothyroidism in the developed world. In younger patients, it is more likely to be associated with goiter; in older patients, the gland may be totally destroyed by the immunologic process, and the only trace of the disease is a persistently positive test for TPO Abs. Similarly, the end stage of Graves disease may be hypothyroidism, occurring spontaneously or following destructive therapy with radioactive iodine or thyroidectomy. Thyroid glands afflicted with autoimmune inflammation are particularly susceptible to excessive iodide intake (eg, ingestion of kelp tablets, iodide-containing cough preparations, or the antiarrhythmic drug amiodarone) or intravenous administration of iodide-containing radiographic contrast media. Large amounts of iodide block thyroid hormone synthesis via the Wolff-Chaikoff effect (see earlier), producing iodine-induced hypothyroidism with goiter in the patient with an abnormal thyroid gland; the normal gland *escapes* from the Wolff-Chaikoff effect or iodide block, but for unclear reasons, autoimmunity renders the gland more sensitive to the inhibitory effects of iodine. Hypothyroidism may occur during the late phase of subacute thyroiditis or silent thyroiditis; this is usually transient, but it may be permanent especially in silent thyroiditis, where permanent hypothyroidism occurs in about 25% of patients. Certain drugs can block hormone synthesis and produce hypothyroidism with goiter; at present, the most common pharmacologic causes of hypothyroidism (other than iodide) are lithium carbonate, used for the treatment of bipolar disease, and amiodarone. Chronic therapy of hyperthyroidism with the antithyroid drugs PTU and methimazole have the same effects. Interferon alfa, used infrequently now to treat hepatitis C and other conditions, can cause altered immunity that can result in hypothyroidism due to Hashimoto thyroiditis. Inborn errors of thyroid hormone synthesis, called thyroid dyshormonogenesis, result from genetic deficiencies in enzymes necessary for hormone biosynthesis. These effects may be complete, resulting in a syndrome of severe congenital hypothyroidism (cretinism) with goiter; or partial, resulting in goiter with milder hypothyroidism. At least five separate biosynthetic abnormalities have been reported: (1) impaired transport of iodine; (2) deficient TPO with impaired oxidation of iodide to iodine and failure to incorporate iodine into TG; (3) impaired coupling of iodinated tyrosines to triiodothyronine or tetraiodothyronine; (4) absence or deficiency of iodotyrosine deiodinase, so that iodine is not conserved within the gland; and (5) excessive production of metabolically inactive iodoprotein by the thyroid gland (see Figure 7–6). The latter may involve impaired or abnormal TG synthesis.

Pituitary and hypothalamic deficiencies as causes of hypothyroidism are quite rare and are usually associated with other symptoms and signs of pituitary insufficiency (see Chapter 4). Peripheral RTH is discussed later.

Pathogenesis

Thyroid hormone deficiency affects virtually every tissue in the body, so that the symptoms are multiple. Pathologically, the most characteristic finding is the accumulation of glycosaminoglycans—mostly hyaluronic acid—in interstitial tissues. Accumulation of this hydrophilic substance and increased capillary permeability to albumin account for the interstitial nonpitting edema that is particularly evident in the skin, heart muscle, and striated muscle. The accumulation is due not to excessive synthesis but to decreased metabolism of glycosaminoglycans.

Clinical Presentations and Findings

A. Newborn infants (cretinism) The term cretinism was originally applied to infants born in areas of severe iodine deficiency with mental retardation, short stature, a characteristic puffy appearance of the face and hands, and (frequently) deaf mutism and neurologic signs of pyramidal and extrapyramidal tract abnormalities. In the United States, neonatal screening programs have revealed that in the Caucasian population the prevalence of sporadic neonatal hypothyroidism is 1:5000, whereas in the African American population the prevalence is only 1:32,000. In areas of iodine sufficiency, neonatal hypothyroidism typically results from failure of the thyroid to descend during embryonic development from its origin at the base of the tongue to its usual site in the lower anterior neck, resulting in an absent or *ectopic thyroid* gland that functions poorly. Placental transfer to the embryo of blocking anti-TSH receptor Abs from a mother with Hashimoto thyroiditis may result in agenesis of the thyroid gland and athyreotic cretinism, but usually it only causes transient hypothyroidism. Inherited defects in thyroid hormone biosynthesis may also cause neonatal hypothyroidism and goiter. Other possible causes of neonatal hypothyroidism include exposure during pregnancy to iodides, antithyroid drugs given to the mother, or inadvertent administration of radioactive iodine for thyrotoxicosis or thyroid cancer.

The signs of hypothyroidism in newborns include respiratory difficulty, cyanosis, jaundice, poor feeding, hoarse cry, umbilical hernia, and marked retardation of bone maturation. The proximal tibial epiphysis and distal femoral epiphysis are present in almost all full-term infants with a body weight over 2500 g. Absence of these epiphyses strongly suggests hypothyroidism. The introduction of routine screening of newborns for TSH or T_4 in the developed world has been a major public health achievement, because early diagnosis can prevent permanent mental retardation. A drop of blood obtained by heel stick 24 to 48 hours after birth is placed on filter paper and sent to a central laboratory. A serum T_4 under 6 µg/dL or a serum TSH over 25 mU/L is suggestive of neonatal hypothyroidism. The diagnosis can then be confirmed by repeat testing and radiologic evidence of retarded bone age. Note that euthyroid infants born to hypothyroid mothers who have been inadequately treated with levothyroxine during pregnancy are not hypothyroid, but may have diminished intellectual potential later in childhood—emphasizing the importance of maintaining the mother in a euthyroid state throughout pregnancy.

B. Children and adolescents   Hypothyroidism in children and adolescents is characterized by retarded growth and short stature, the typical signs and symptoms of hypothyroidism seen in adults (see later), and variable but usually declining school performance. Precocious puberty may occur, and there may be enlargement of the sella turcica due to pituitary thyrotrope hyperplasia associated with augmented TSH production.

C. Adults   In adults, the common features of moderate to severe hypothyroidism include easy fatigability; cold sensitivity; weight gain (generally <10-20 lb); constipation; menstrual abnormalities, especially menorrhagia; and muscle cramps. Physical findings may include a cool, rough, dry skin; puffy face and hands; a hoarse, husky voice; and slow reflexes. Reduced

conversion of carotene to vitamin A and increased blood levels of carotene may give the skin a yellowish color. However, many or all of the symptoms and signs are diminished or absent in patients with milder degrees of thyroid failure.

1. **Cardiovascular signs**—Hypothyroidism is manifested by impaired ventricular contraction, bradycardia, and increased peripheral resistance, resulting in diminished cardiac output. The electrocardiogram (ECG) may reveal low voltage of QRS complexes and P and T waves, with improvement in response to therapy. Cardiac enlargement may occur, due in part to interstitial edema, nonspecific myofibrillary swelling, and left ventricular dilation, as well as nonhemodynamically significant pericardial effusion (Figure 7–29). The degree of pericardial
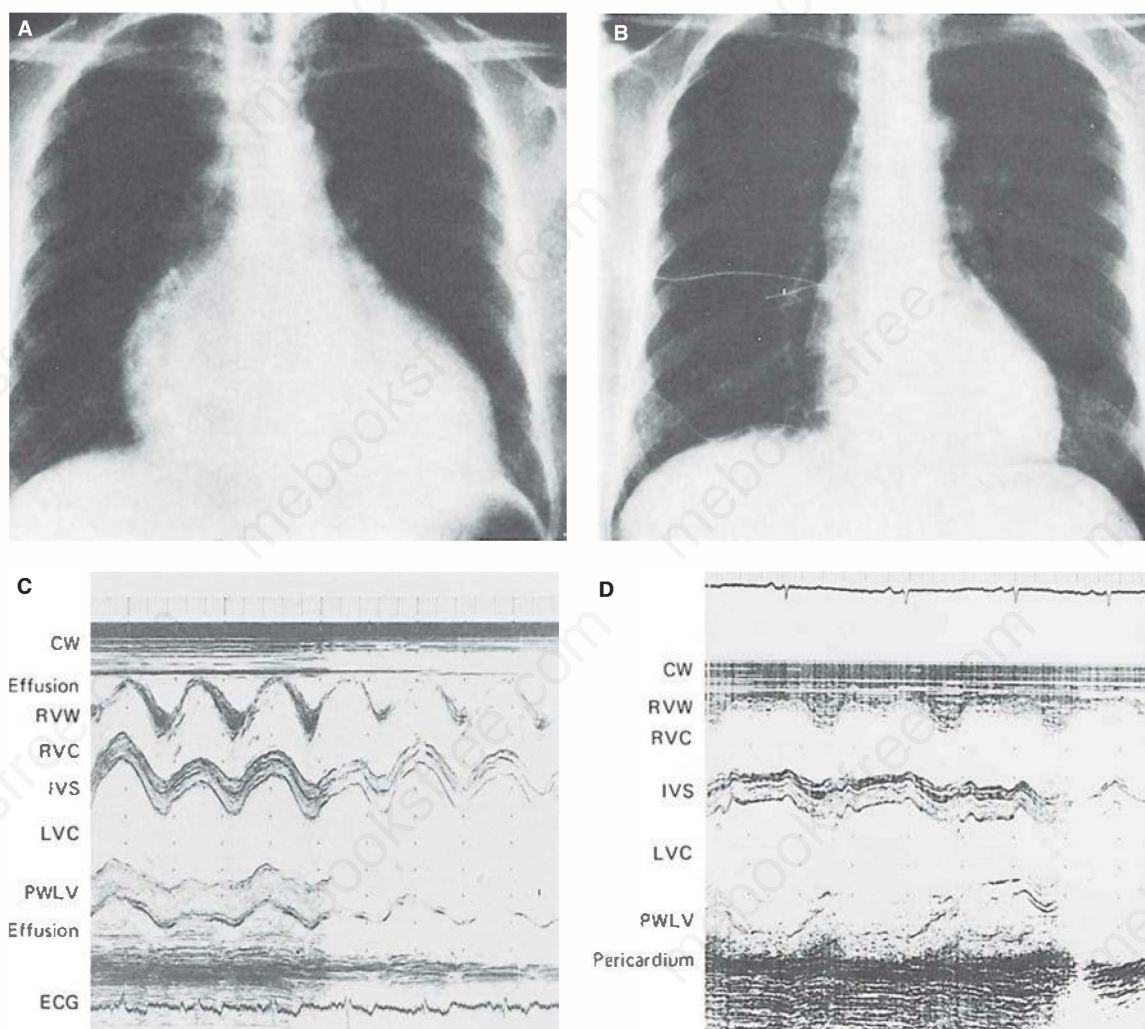
FIGURE 7–29   **Top.** Chest radiograph studies of patient with hypothyroid cardiomyopathy. **A.** Before therapy, showing pronounced cardio-megaly. **B.** Six months after institution of thyroxine therapy, the heart size has returned to normal. (Reproduced with permission from Reza MJ, Abbasi AS. Congestive cardiomyopathy in hypothyroidism. *West J Med*. 1975 Sep;123(3):228-230.)
Bottom. Echocardiogram of a 29-year-old woman with hypothyroidism **C.** before and **D.** after 2 months of therapy with levothyroxine sodium. Note disappearance of pericardial effusion following levothyroxine therapy. (CW, chest wall; RVW, right ventricular wall; RVC, right ventricular cavity; IVS, inter-ventricular septum; LVC, left ventricular cavity; PWLV, posterior wall left ventricle.) (Reproduced with permission from Sokolow M, McIlroy MB. *Clinical Cardiology*, 4th ed. New York; McGraw-Hill Education; 1986.)

effusion can easily be determined by echocardiography. Although cardiac output is reduced, congestive heart failure and pulmonary edema are rarely noted. There is controversy about whether hypothyroidism induces coronary artery disease, but coronary artery disease is more common in patients with hypothyroidism, likely related to increased levels of total cholesterol, LDL cholesterol, and possibly other nontraditional atherogenic factors such as lipoprotein A and homocysteine. In patients with angina pectoris, hypothyroidism may protect the heart from ischemic stress, and replacement therapy may aggravate the angina by increasing myocardial oxygen consumption.

2. **Pulmonary function**—In the adult, hypothyroidism is characterized by shallow, slow respirations and impaired ventilatory responses to hypercapnia or hypoxia. Respiratory failure is a major problem in patients with myxedema coma (see later).

3. **Intestinal peristalsis**—Peristalsis is markedly slowed, resulting in chronic constipation and occasionally severe fecal impaction or ileus.

4. **Renal function**—Renal function is impaired, with decreased glomerular filtration rate and impaired ability to excrete a water load. This predisposes the hypothyroid patient to hyponatremia from water intoxication if excessive free water is administered.

5. **Anemia**—There are at least four mechanisms that may contribute to anemia in patients with hypothyroidism: (1) impaired hemoglobin synthesis as a result of T_4 deficiency; (2) iron deficiency from increased iron loss with menorrhagia, as well as impaired intestinal absorption of iron; (3) folate deficiency from impaired intestinal absorption of folic acid; and (4) pernicious anemia, with vitamin B_{12}-deficient megaloblastic anemia. Pernicious anemia is often part of a cluster of autoimmune diseases, including hypothyroidism due to Hashimoto or autoimmune thyroiditis associated with thyroid autoantibodies, pernicious anemia associated with parietal cell autoantibodies, diabetes mellitus associated with islet cell autoantibodies, and adrenal insufficiency associated with adrenal autoantibodies (Schmidt syndrome; see Chapter 2). These conditions are part of what has been termed the **polyendocrine failure syndrome**.

6. **Neuromuscular system**—Many patients complain of symptoms referable to the neuromuscular system, including severe muscle cramps, paresthesias, and muscle weakness.

7. **Central nervous system**—Symptoms may include chronic fatigue, lethargy, and inability to concentrate. Patients with hypothyroidism are usually quite placid but can be severely depressed or even extremely agitated (myxedema madness).

8. **Reproductive system**—Hypothyroidism impairs the conversion of estrogen precursors to estrogens resulting in altered FSH and LH secretion and in anovulatory cycles and infertility. This may be associated with menorrhagia, which may also be due, in part, to altered platelet function. Men may have decreased libido and erectile dysfunction.

Diagnosis

The combination of a low serum FT_4 and an elevated serum TSH is diagnostic of primary hypothyroidism (Figure 7–30). Serum T_3 levels are variable and may be within the normal range. Commonly, the serum FT_4 is normal or low-normal, and the serum TSH is elevated slightly, a situation termed **subclinical hypothyroidism**. This represents the mildest form of hypothyroidism and is a consequence of the very sensitive feedback relationship between the thyroid and the pituitary gland. In this situation, a small decrement in thyroid hormone output by the thyroid gland, in which serum T_4 levels are still within the normal range, results in a serum TSH level that is elevated, albeit usually less than 10 mU/L. Subclinical hypothyroidism is usually due to underlying Hashimoto thyroiditis, which can be confirmed by assessing anti-TPO Ab titers. In patients with secondary or central hypothyroidism, the serum FT_4 will be low and serum TSH will be low or normal, rather than elevated. Since serum TSH rises with advancing age, some healthy elderly persons may have mildly elevated serum TSH (eg, 5-7 mU/L), that can mimic subclinical hypothyroidism.

The patient may be taking thyroid hormone when first seen. A palpable or enlarged thyroid gland and a positive test for thyroid autoantibodies would suggest underlying Hashimoto thyroiditis, in which case the medication should be continued. If antithyroid Abs are absent and the indication for therapy uncertain, the medication could be withdrawn for 6 weeks and determination made for FT_4 and TSH. The 6-week period of withdrawal is necessary because of the long half-life of T_4 (7 days) and to allow the pituitary gland to recover after a long period of suppression. The pattern of recovery of thyroid function after withdrawal of T_4 is noted in Figure 7–31. In hypothyroid individuals, serum TSH becomes elevated at 5 to 6 weeks and FT_4 remains subnormal, whereas both are normal after 6 weeks in euthyroid individuals.

The clinical picture of fully developed myxedema is usually quite clear, but the symptoms and signs of mild or subclinical
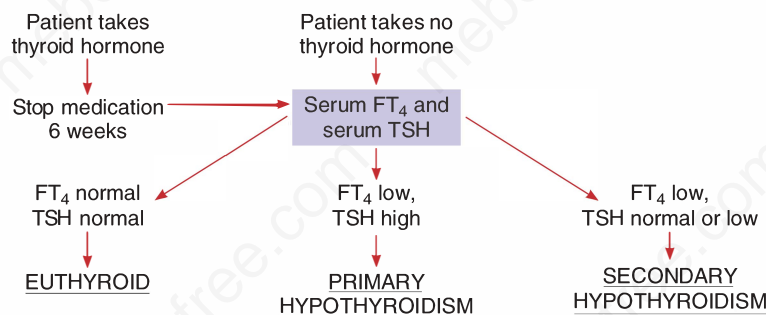
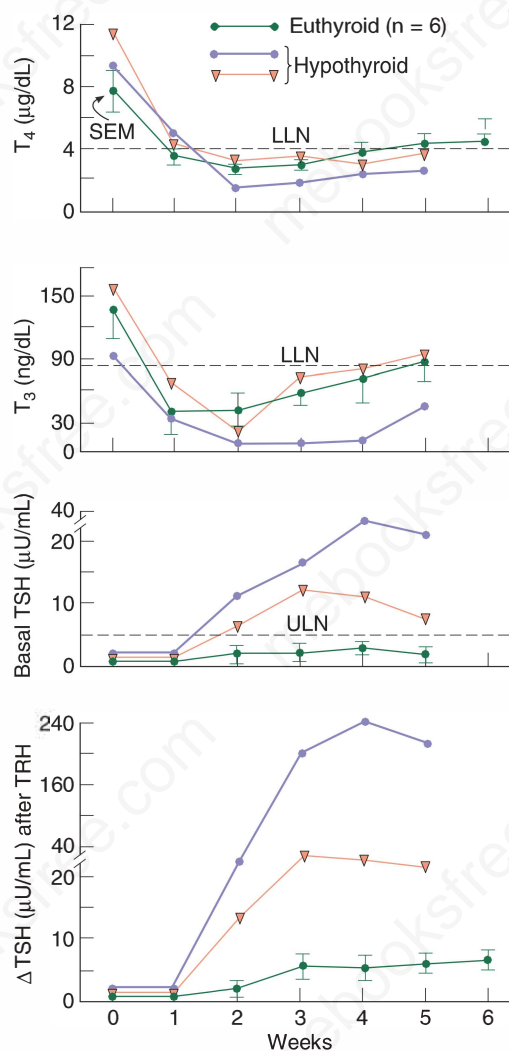FIGURE 7–30 Diagnosis of hypothyroidism.

FIGURE 7–31 Changes in T_4, T_3, TSH, and TRH response following abrupt withdrawal of suppressive thyroxine therapy. Note that in euthyroid individuals, the T_4 may not return to normal until 6 weeks after withdrawal of therapy and that serum TSH is never elevated. In hypothyroid patients, TSH may be elevated as early as 2 weeks after withdrawal of therapy, and TRH response is exaggerated. (LLN, lower limit of normal; ULN, upper limit of normal.) (Adapted with permission from Vagenakis AG, Braverman LE, Azizi F, et al. Recovery of pituitary thyrotropic function after withdrawal of prolonged thyroid suppression therapy. *N Engl J Med*. 1975 Oct 2;293(14):681-684.)

hypothyroidism may be very subtle or absent. This has led to the recommendation by some professional organizations that screening for hypothyroidism be undertaken, especially in high-risk groups, such as older women where the prevalence is high (up to 20% in women age >65), and in pregnant women, where untreated hypothyroidism may cause adverse outcomes in the child.

At times, patients with hypothyroidism present with unusual features: neurasthenia with symptoms of muscle cramps, paresthesias, and weakness; refractory anemia; disturbances in

reproductive function, including infertility, delayed or precocious puberty, or menorrhagia; idiopathic edema or pleuropericardial effusions; retarded growth; obstipation; chronic rhinitis or hoarseness due to edema of nasal mucosa or vocal cords; and severe depression progressing to emotional instability or even frank paranoid psychosis. In the elderly, hypothyroidism may present with apathy and withdrawal, often attributed to senility. In such cases, assessing thyroid function with serum FT_4 and TSH measurements confirms or rules out hypothyroidism as a contributing factor.

Complications

A. Myxedema coma   Myxedema coma, an extremely rare condition, is the end stage of untreated hypothyroidism (see also Chapter 24). It is characterized by progressive weakness, stupor, hypothermia, hypoventilation, hypoglycemia, and hyponatremia, and it may ultimately result in shock and death. It occurs most frequently in the winter in older female patients with underlying pulmonary and vascular disease, and the mortality rate may be greater than 50%.

The patient (or a family member if the patient is comatose) may recall previous thyroid disease, radioiodine or radiation therapy to the neck area, or thyroidectomy. The medical history is one of gradual onset of lethargy progressing to stupor or coma. Examination reveals bradycardia and marked hypothermia, with body temperature as low as 24°C (75°F). The patient is usually an obese elderly woman with yellowish skin, a hoarse voice, a large tongue, thin hair, puffy eyes, ileus, and slow reflexes. There may be signs of other illnesses such as pneumonia, myocardial infarction, cerebral thrombosis, or gastrointestinal bleeding. Seizures, bleeding episodes, hypocalcemia, or hypercalcemia may occur. Laboratory clues to the diagnosis of myxedema coma include lactescent serum, high serum carotene, hyponatremia, elevated BUN and creatinine, elevated serum cholesterol, and increased cerebrospinal fluid protein. Pleural, pericardial, or abdominal effusions with high protein content may be present. Serum tests reveal a low FT_4 and a markedly elevated TSH. Thyroid autoantibodies are usually strongly positive, indicating underlying Hashimoto thyroiditis. The ECG shows sinus bradycardia and low voltage. If laboratory studies are not immediately available, which is frequently the case, the diagnosis must be made clinically.

The pathophysiology of myxedema coma involves three major features: (1) CO_2 retention and hypoxia, (2) fluid and electrolyte imbalance, and (3) hypothermia. CO_2 retention and hypoxia are probably due in large part to a marked depression in the ventilatory responses to hypoxia and hypercapnia, although factors such as obesity, heart failure, ileus, immobilization, pneumonia, pleural or peritoneal effusions, central nervous system depression, and weak chest muscles may also contribute. Impairment of ventilatory drive is often severe, and assisted respiration is almost always necessary in patients with myxedema coma. Thyroid hormone therapy in patients with myxedema corrects the hypothermia and improves the ventilatory response to hypoxia. Water intoxication, due to reduced renal perfusion and impaired free water clearance, is the major fluid and electrolyte disturbance. This causes

hyponatremia and is best managed by free water restriction. Hypothermia may not be recognized if the ordinary clinical thermometer only goes down to about 34°C (93°F); and a thermometer that registers a broader scale must be used to obtain accurate body temperature readings. Active rewarming of the body is contraindicated, because it may induce vasodilation and vascular collapse. A rise in body temperature is a useful indication of therapeutic effectiveness of T_4 therapy.

Disorders that may precipitate myxedema coma include heart failure, pneumonia, or administration of sedative or narcotic drugs to a patient with severe hypothyroidism. Adrenal insufficiency may occur occasionally in association with myxedema coma, either due to functional pituitary impairment or concurrent autoimmune adrenal insufficiency (Schmidt syndrome). Concomitant glucocorticoid therapy is recommended until normal adrenal function can be documented by laboratory testing. It is important to differentiate myxedema coma due to primary thyroid gland failure from that due to pituitary failure (central hypothyroidism). In the latter situation, glucocorticoid replacement is essential. Clinical clues to the presence of pituitary disease include the following: a history of amenorrhea or impotence, scanty pubic or axillary hair, and normal serum cholesterol and normal or low serum TSH levels. CT or MRI may reveal an enlarged sella turcica. The treatment of myxedema coma is discussed later.

B. Hypothyroidism and heart disease   In the past, treatment of patients with severe hypothyroidism and heart disease, particularly coronary artery disease, was very difficult, because levothyroxine replacement was sometimes associated with exacerbation of angina, heart failure, or myocardial infarction. Now that β-adrenergic–blocking agents, coronary stenting, and coronary artery bypass surgery are available, hypothyroid patients with coronary artery disease can be treated medically or surgically first, and T_4 replacement therapy is then better tolerated.

C. Hypothyroidism and neuropsychiatric disease
Hypothyroidism is often associated with depression, which may be quite severe. More rarely, severely hypothyroid patients may become confused, paranoid, or even manic (myxedema madness). Screening of psychiatric admissions with serum FT_4 and TSH is an efficient way to identify these patients, who frequently respond to T_4 therapy alone or in combination with psychopharmacologic agents. The effectiveness of T_4 therapy in depressed hypothyroid patients has given rise to the hypothesis that the addition of T_3 or T_4 to psychotherapeutic regimens for depression may be helpful in patients without demonstrable thyroid disease, but this concept has been difficult to prove.

Treatment

A. Treatment of hypothyroidism   Hypothyroidism is treated with T_4, which is available in pure form and is stable and inexpensive. Because T_4 is converted to T_3 in peripheral tissues, both hormones become available, even though only one is administered. Desiccated thyroid is now considered obsolete because it

contains both T_4 and T_3, and triiodothyronine (as liothyronine) is unsatisfactory because of its rapid absorption, short half-life, and transient effects. The half-life of T_4 is about 7 days, so it needs to be given only once daily. It is well absorbed, and blood levels are easily monitored by following FT_4 and serum TSH levels. T_4 and T_3 levels are stable throughout the day, although there is a slight increase in free T_4 levels several hours after the medication is taken; this is not clinically significant.

B. Dosage of levothyroxine   Replacement doses of T_4 in adults range from 0.05 to 0.2 mg/d, with a mean of about 0.125 mg/d. The dose of T_4 varies according to the patient's age and body weight (Table 7–7). Because of more rapid T_4 metabolism, infants and young children require a surprisingly high dose of T_4 compared with adults. In adults, the mean replacement dose of T_4 is about 1.7 μg/kg/d, or about 0.8 μg/lb/d. In older adults, the replacement dose is usually lower, about 1.6 μg/kg/d, or about 0.7 μg/lb/d. For TSH suppression in patients postthyroidectomy for thyroid cancer, the average dose of T_4 is about 2.2 μg/kg/d (1 μg/lb/d). In most patients with hypothyroidism, one can begin treatment with the full estimated dose requirement. After 4 to 6 weeks, the final dose is adjusted based on the serum TSH level. The goal is to normalize the serum TSH, which is typically between 0.5 and 4 mU/L. In older patients or patients with underlying heart disease, it is best to start with a low dose of T_4 (eg, 0.025 mg/d) and increase the dose at 4- to 6-week intervals based on serum TSH measurements. Malabsorptive states or concurrent administration of soy products, aluminum hydroxide antacids, bile acid–binding resins such as cholestyramine and colestipol, calcium supplements, sucralfate, or iron compounds decrease T_4 absorption. In these patients, T_4 should be given before breakfast, when the stomach is empty, and the other compounds taken 4 hours later. Other drugs, such as the antiseizure medication carbamazepine, may increase thyroid hormone requirements by increasing T_4 metabolism. Proton pump inhibitors such as omeprazole may also impair T_4 absorption, possibly because gastric acid is needed for dissolution of the tablets. Estrogen replacement may increase T_4 requirements by increased binding of free T_4 to TBG. T_4 has a sufficiently long half-life (7 days), so that if the patient is unable

TABLE 7–7 Replacement doses of levothyroxine.

Age	Dose of Levothyroxine (μg/kg/d)
0-6 mo	10-15
7-11 mo	6-8
1-5 y	5-6
6-10 y	4-5
11-20 y	1-3
Adult	1-2

Data from Braverman LE, Utiger RD. *Werner and Ingbar's The Thyroid*, 7th ed. Philadelphia: Lippincott; 1996.

to take medications by mouth for a few days, omitting T_4 therapy will not be detrimental. If a hospitalized patient is being managed by sustained parenteral therapy, the parenteral dose of T_4 is about 75% to 80% of the usual oral dose.

C. Treatment of myxedema coma   Myxedema coma is an acute medical emergency and should be treated in the intensive care unit. Blood gases must be monitored regularly, and the patient usually requires intubation and mechanical ventilation. Associated illnesses such as infections or heart failure must be sought and appropriately treated. Intravenous fluids should be administered with caution, and excessive free water intake must be avoided. Because patients with myxedema coma may absorb drugs poorly, it is imperative to give T_4 intravenously. These patients have marked total body depletion, and they should receive an initial loading dose of 300 to 400 μg of T_4 intravenously followed by 80% of the calculated full replacement dose intravenously daily. Management of patients with known or suspected heart disease is described later. If, after a few days, the clinical response has been suboptimal, some experts recommend adding intravenous T_3 to the regimen at a dose of 5 μg every 6 hours. The clinical guides to improvement are a rise in body temperature and the return of normal cerebral and respiratory function. The possibility of concomitant adrenal insufficiency (due to autoimmune adrenal disease or pituitary insufficiency) needs to be considered and ruled out with a cosyntropin stimulation test (see Chapter 9). Full adrenal support should then be administered (eg, hydrocortisone hemisuccinate) 100 mg intravenously, followed by 50 mg intravenously every 6 hours, tapering the dose over 7 days. Adrenal support can be withdrawn if the pretreatment plasma cortisol is 20 μg/dL or greater or if results of a cosyntropin stimulation test are within normal limits. When giving T_4 intravenously in large doses, there is an inherent risk of precipitating angina, heart failure, or arrhythmias in older patients with underlying coronary artery disease.

At one time, the mortality rate of myxedema coma was about 80%. The prognosis has been vastly improved as a result of recognition of the importance of mechanically assisted ventilation and the use of intravenous T_4. At present, the outcome probably depends on how well the underlying comorbidities and consequences of severe hypothyroidism can be managed.

D. Severe hypothyroidism with heart disease   In older patients—particularly those with known cardiovascular disease—it is wise to start treatment slowly. T_4 is given in a dosage of 0.025 mg/d for 2 to 4 weeks, increasing by 0.025 mg every 2 to 4 weeks until a daily dose of 0.075 mg is reached. This dose is continued for about 6 weeks. Serum TSH is then measured and the dosage adjusted accordingly. It typically takes 2 to 4 months for a patient to come into equilibrium on full dosage. In some patients, the heart is very sensitive to the level of circulating T_4, and if angina pectoris or cardiac arrhythmia develops, it is essential to reduce the dose of T_4 immediately.

Adverse Effects of T_4 Therapy

There are no reported instances of allergy to pure T_4, although it is possible that a patient may develop an allergy to the coloring dye or some component of the tablet. The major toxic effects of T_4 overdosage are symptoms and signs of hyperthyroidism. Palpitations are the most common thyrotoxic cardiac symptom, and arrhythmias, particularly paroxysmal atrial tachycardia or atrial fibrillation, may occur. Insomnia, tremor, restlessness, and excessive warmth may also be troublesome. Simply omitting the daily dose of T_4 for 3 days and then reducing the dosage will correct the problem.

Increased bone resorption and osteoporosis have been associated with longstanding hyperthyroidism and will also develop in postmenopausal women chronically overtreated with T_4. This can be prevented by regular monitoring and by maintaining normal serum TSH levels in patients receiving long-term replacement therapy. In patients receiving TSH-suppressive therapy for thyroid cancer, if serum FT_4 levels are kept in the upper range of normal—even if TSH is suppressed—the adverse effects of T_4 therapy on bone will be minimal. In addition, concomitant administration of estrogen or bisphosphonates to postmenopausal women receiving high-dose T_4 therapy minimizes bone resorption.

Course and Prognosis

The course of untreated hypothyroidism is one of slow deterioration, potentially leading eventually to myxedema coma and death. With appropriate treatment, however, the long-term prognosis is excellent. Because of the long half-life (7 days) of T_4, it takes time to establish equilibrium on a fixed dose. Therefore, it is important to monitor the serum FT_4 and TSH every 4 to 6 weeks until equilibrium is reached. Thereafter, FT_4 and TSH can be monitored once in every 6 to 12 months. The dose of T_4 is increased by about 50% in most women during pregnancy with the greatest dosage increases in those women with the least thyroid reserve (ie, in women who have had a thyroidectomy or who have received radioactive iodine therapy). Older patients metabolize T_4 more slowly, and the dose likely gradually decreases with age. A small proportion of patients with hypothyroidism treated with levothyroxine monotherapy continue to have symptoms consistent with thyroid hormone deficiency, despite having normal serum FT_4 and TSH levels. The cause of this phenomenon is not known, but some experts have recommended a trial of combination T_4 and T_3 therapy in selected patients, despite the fact that controlled trials that have compared T_4 monotherapy to T_4 plus T_3 combination therapy showed no apparent benefit.

HYPERTHYROIDISM AND THYROTOXICOSIS

Thyrotoxicosis is the clinical syndrome that results when tissues are exposed to high levels of circulating thyroid hormones. It results in a generalized acceleration of metabolic processes. In most instances, thyrotoxicosis is due to hyperactivity of the thyroid gland, or hyperthyroidism. Occasionally, thyrotoxicosis may be due to other

TABLE 7–8 Conditions associated with thyrotoxicosis.

1. Diffuse toxic goiter (Graves disease)
2. Toxic adenoma (Plummer disease)
3. Toxic multinodular goiter
4. Subacute thyroiditis
5. "Silent" thyroiditis
6. Thyrotoxicosis factitia
7. Rare forms of thyrotoxicosis: ovarian struma, metastatic thyroid carcinoma (follicular), hydatidiform mole, hamburger thyrotoxicosis TSH-secreting pituitary tumor, pituitary resistance to T_3 and T_4

causes such as excessive ingestion of thyroid hormone or, very rarely, excessive secretion of thyroid hormones from an ovarian tumor (struma ovarii). Some experts reserve the term "thyrotoxicosis" for all conditions in which serum levels of thyroid hormones are elevated, while the word "hyperthyroidism" is reserved for those conditions in which the thyroid gland is synthesizing and secreting excessive quantities of thyroid hormones. Thus, Graves disease, toxic multinodular goiter, and TSH-secreting tumor are examples of "hyperthyroidism," while thyroiditis and factitious thyrotoxicosis are examples of thyrotoxicosis, but not true "hyperthyroidism." The various forms of thyrotoxicosis are listed in Table 7–8. These syndromes will be discussed individually later.

1. DIFFUSE TOXIC GOITER (GRAVES DISEASE)

Graves disease is the most common form of thyrotoxicosis. Females are involved about 5 times more commonly than males. The disease may occur at any age, with a peak incidence in the 20- to 40-year age group (see section Thyroid Autoimmunity, earlier). The syndrome consists of one or more of the following features: (1) thyrotoxicosis, (2) goiter, (3) ophthalmopathy (exophthalmos), and (4) dermopathy (pretibial myxedema).

Etiology

Graves disease is currently viewed as an autoimmune disease of unknown cause. There is a strong familial predisposition, in that about 15% of patients with Graves disease have a close relative with the same disorder, and about 50% of relatives of patients with Graves disease have circulating thyroid autoantibodies. There is a much higher concordance of Graves disease in monozygotic twins compared to dizygotic twins, but because the concordance rate in monozygotic twins is far less than 100%, there must be environmental factors that come into play as well. Proposed environmental triggers include stress, tobacco use, infection, and iodine exposure. The postpartum state, which may be associated with heightened immune function, also may trigger the development of Graves disease in genetically susceptible women.

Pathogenesis

In Graves disease, T lymphocytes become sensitized to antigens within the thyroid gland and stimulate B lymphocytes
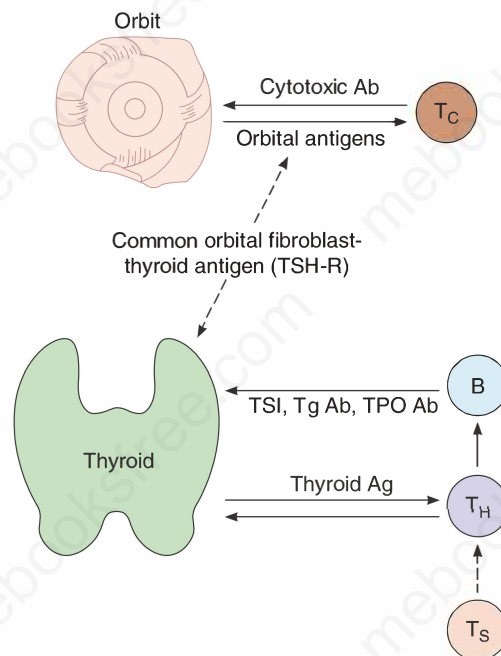

FIGURE 7–32 One theory of the pathogenesis of Graves disease. There is a defect in suppressor T lymphocytes (T_s) that allows helper T lymphocytes (T_H) to stimulate B lymphocytes (B) to synthesize thyroid autoantibodies. The thyroid-stimulating immunoglobulin (TSI) is the driving force for thyrotoxicosis. The inflammatory process in the orbital muscles may be due to sensitization of cytotoxic T lymphocytes (T_c), or killer cells, to orbital antigens in association with cytotoxic antibodies. The thyroid and the eye are linked by a common antigen, the TSH-R, found in thyroid follicular cells and orbital fibroblasts. It is not yet clear what triggers this immunologic cascade. (Tg Ab, thyroglobulin antibody; TPO Ab, thyroperoxidase or microsomal antibody; Ag, antigen; Ab, antibody.)

to synthesize Abs to these antigens (see Chapter 2; see also the section Thyroid Autoimmunity earlier and Figure 7–32). One such Ab is directed against the TSH receptor in the thyroid cell membrane, stimulating thyroid gland growth and function. The Ab is called a thyroid-stimulating Ab (TSAb), or TSI. The presence of this circulating Ab is positively correlated with active disease and with relapse of the disease following therapy with antithyroid drugs. There is an underlying genetic predisposition, but it is not clear what triggers the initial onset of hyperthyroidism. Some factors that may incite the immune response of Graves disease are (1) pregnancy, particularly the postpartum period; (2) iodide excess, particularly in geographic areas of iodide deficiency, where the lack of iodide may hold latent Graves disease in check; (3) interferon alfa, perhaps by modifying immune responsiveness; (4) viral or bacterial infections; and (5) psychological stress. The pathogenesis of ophthalmopathy may involve cytotoxic lymphocytes (killer cells) and cytotoxic Abs sensitized to a common antigen such as the TSH receptor that is found in orbital fibroblasts, orbital muscle, and thyroid tissue (see Figure 7–32). Cytokines from these sensitized lymphocytes may cause activation and proliferation of orbital fibroblasts and preadipocytes, resulting in

FIGURE 7–33 Dermopathy of Graves disease. Marked thickening of the skin is noted, usually over the pretibial area. Thickening occasionally extends downward over the ankle and the dorsal aspect of the foot but almost never above the knee.

increased amounts of retro-orbital fat and glycosaminoglycans, as well as swollen extraocular muscles; this results in proptosis (protrusion) of the globes and diplopia, as well as redness, congestion, and conjunctival and periorbital edema (thyroid-associated ophthalmopathy or orbitopathy). For unclear reasons, smoking cigarettes is a potent risk factor for the development of thyroid-associated ophthalmopathy, possibly related to effects of tobacco smoke on orbital adipogenesis.

The pathogenesis of thyroid dermopathy (pretibial myxedema) (Figure 7–33) and the rare subperiosteal inflammation on the phalanges of the hands and feet (thyroid osteopathy or acropachy) (Figure 7–34) may also involve lymphocyte cytokine stimulation of fibroblasts in these locations. Many symptoms of thyrotoxicosis suggest a state of catecholamine excess, including tachycardia, tremor, sweating, lid lag, and stare. However, circulating levels of epinephrine and norepinephrine are normal or low; thus, in Graves disease, the body appears to have an increased sensitivity to catecholamines. This may be due in part to a thyroid hormone–mediated increase in membrane-bound β-adrenergic receptors in various tissues.

Clinical Features

A. Symptoms and signs   In younger individuals, common manifestations include palpitations, nervousness, easy fatigability, hyperkinesia, diarrhea, excessive sweating, intolerance to heat, and preference for cold. There is often marked weight loss without loss of appetite. Thyroid enlargement, thyrotoxic eye

FIGURE 7–34 Radiograph of hand of a patient with thyroid osteopathy. Note marked periosteal thickening of the proximal phalanges.

signs (see later), and mild tachycardia commonly occur. Muscle weakness and loss of muscle mass may be so severe that the patient cannot rise from a chair without assistance. In children, rapid growth with accelerated bone maturation occurs. In patients over age 60, cardiovascular and myopathic manifestations predominate; the most common presenting complaints are palpitations, dyspnea on exertion, tremor, nervousness, and weight loss (see Chapter 23).

The eye signs of Graves disease have been classified by Werner as set forth in Table 7–9. This classification is useful in

TABLE 7–9 Classification of eye changes in Graves' disease.

Class	Definition
0	No signs or symptoms
1	Only signs, no symptoms (signs limited to upper lid retraction, stare, lid lag)
2	Soft tissue involvement (symptoms and signs)
3	Proptosis (measured with Hertel exophthalmometer)[a]
4	Extraocular muscle involvement
5	Corneal involvement
6	Sight loss (optic nerve involvement)

[a]Upper limits of normal according to race: Asian,18 mm; white, 20 mm; black, 22 mm. Increase in proptosis of 3-4 mm is mild involvement; 5-7 mm, moderate involvement; and over 8 mm, severe involvement. Other classes can be similarly graded as mild, moderate, or severe.

Reproduced with permission from Werner SC. Classification of the eye changes of Graves' disease. *Am J Ophthalmol*. 1969 Oct;68(4):646-648.

FIGURE 7–35 Hertel exophthalmometer.

FIGURE 7–36 Orbital CT scan in a patient with severe ophthalmopathy and visual failure. Note the marked enlargement of extraocular muscles (X) posteriorly, with compression of the optic nerve at the apex of the orbital cone.

describing the extent of the eye involvement. However, it is not helpful in following the progress of the illness, because one class does not necessarily progress into the next. The first letters of each class form the mnemonic *NO SPECS*. Class 1 involves retraction of the upper lids associated with active thyrotoxicosis and usually resolves spontaneously when the thyrotoxicosis is adequately controlled. Lid retraction can be seen in any form of thyrotoxicosis, because it is due to adrenergic stimulation of the upper eyelid. Classes 2 to 6 represent true infiltrative disease involving orbital muscles and orbital tissues and are specific to Graves disease. Class 2 is characterized by soft tissue involvement with periorbital edema, congestion or redness of the conjunctiva, and edema of the conjunctiva (chemosis). Class 3 consists of proptosis as measured by the Hertel exophthalmometer. This instrument consists of two prisms with a scale mounted on a bar. The prisms are placed on the lateral orbital ridges, and the distance from the orbital ridge to the anterior cornea is measured on the scale (Figure 7–35). The upper limits of normal according to race are listed in the footnote to Table 7–9. Class 4 consists of extraocular muscle involvement, which typically is due to fibrosis and failure of muscle relaxation, limiting the function of the antagonist muscle. The inferior rectus is the muscle most commonly involved in the infiltrative process. By failing to relax normally, upward gaze limitation is the most common physical finding in patients with eye muscle involvement. The muscle next most commonly involved is the medial rectus, impairing lateral gaze. Class 5 is characterized by corneal involvement (keratitis) due to inability to close the eyes completely. Class 6 is a loss of vision from optic nerve involvement, likely due to ischemia of the nerve from compression of the surrounding enlarged extraocular muscles. As noted earlier, thyroid ophthalmopathy is due to infiltration of orbital fibroblasts and the extraocular muscles with lymphocytes and edema fluid due to an inflammatory reaction. The orbit is a cone enclosed by bone, and swelling of the extraocular muscles within this closed space causes proptosis of the globe and impaired muscle movement, resulting in diplopia. Ocular muscle enlargement can be demonstrated by orbital ultrasound, orbital CT scanning, or MRI (Figure 7–36). Although only about one-third of patients have eye involvement clinically, enlarged muscles can be detected by imaging in over 90% of patients.

Thyroid dermopathy consists of thickening of the skin, particularly over the lower tibia, due to accumulation of glycosaminoglycans (see Figure 7–33). It is relatively rare, occurring in about 2% to 3% of patients with Graves disease. It is usually associated with significant ophthalmopathy and with a very high serum titer of TSAb. The skin is markedly thickened with
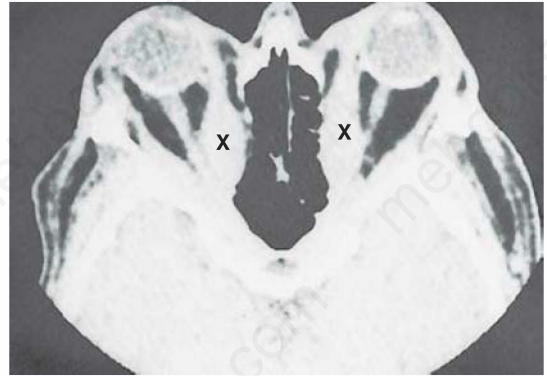
a *peau d'orange* surface, and cannot be picked up between the fingers. Sometimes the dermopathy involves the entire lower leg and may extend onto the feet. Finally, bony involvement (osteopathy or thyroid acropathy), with subperiosteal bone formation and swelling, is particularly evident in the metacarpal bones (see Figure 7–34) and as clubbing of the digits. This too is a relatively rare finding. A more common finding in Graves disease is separation of the fingernails from their beds (onycholysis or Plummer nails), likely caused by rapid growth of the nails (Figure 7–37).

B. Laboratory findings   The laboratory findings in hyperthyroidism are summarized in Figure 7–38. The combination of an elevated FT_4 and a suppressed TSH makes the diagnosis of hyperthyroidism. Approximately 5% of patients have normal FT_4 levels but elevated serum T_3 levels, a situation termed **T_3 thyrotoxicosis**. Also, very mild Graves disease may result in FT_4

FIGURE 7–37 Onycholysis (separation of the nail from its bed) in Graves disease usually resolves spontaneously as the patient improves.

```
                                    FT₄ and TSH

          FT₄ ↑, TSH ↓              FT₄ ↑, TSH ↑      FT₄ N, TSH ↓

        Hyperthyroidism

   Eye signs   No eye signs   No eye signs   TSH-secreting        T₃
   Goiter      No goiter or   uni- or        pituitary tumor
               small goiter   multinodular   GRTH
                              goiter         PRTH
                                                           High    Low    Normal

   Graves      ¹²³I uptake    Toxic adenoma   Early Graves   Euthyroid-sick   Subclinical hyper-
   disease                    or toxic        disease        syndrome         thyroidism usually
                              multinodular    Toxic nodular  Drugs:           due to levothyroxine
               High    Low    goiter          goiter          dopamine,       treatment, mild
                                                               corticosteroids Graves disease,
                                                                               a mild toxic multinodular
   Graves          Spontaneously resolving hyperthyroidism:                   goiter
   disease            Subacute thyroiditis
   Choriocarcinoma    "Silent" thyroiditis
                   Graves disease or toxic nodular goiter in iodine-loaded patient
                   Levothyroxine treatment
                   Rare: Struma ovarii
```
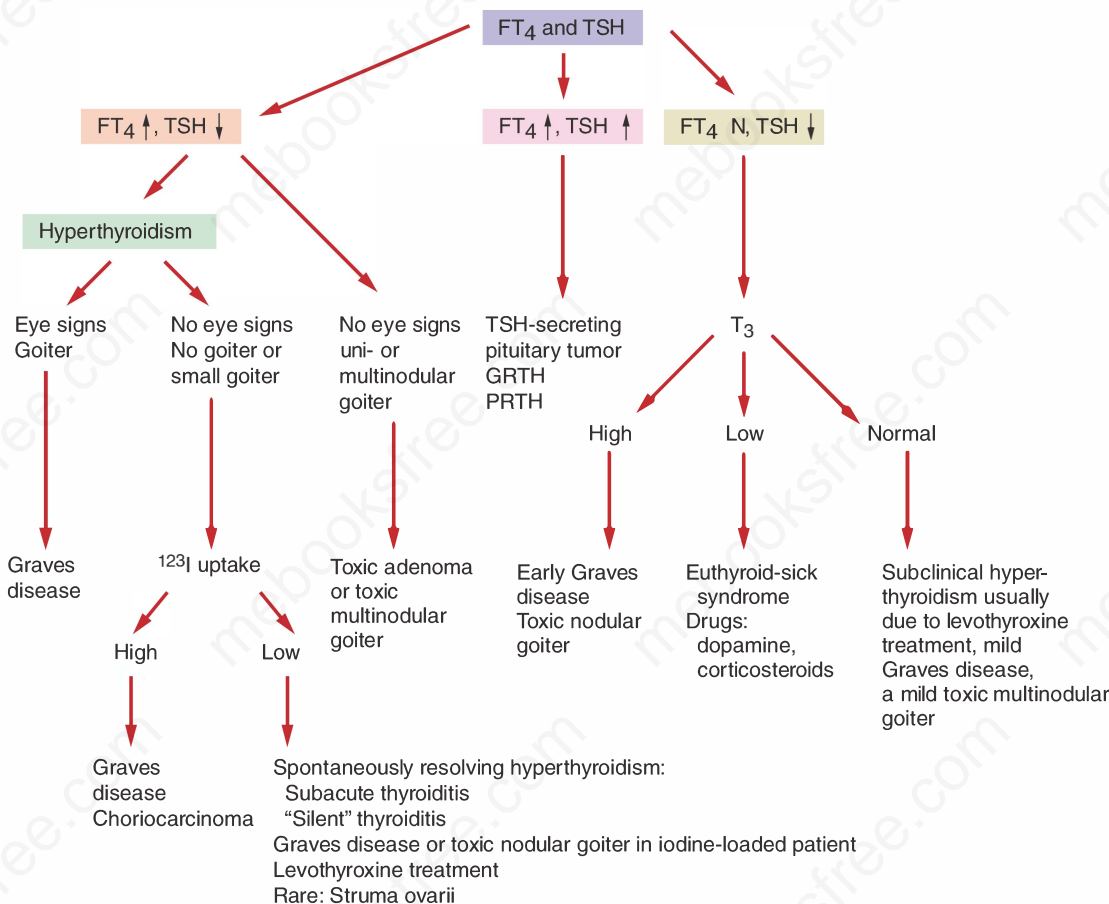
FIGURE 7–38 Laboratory tests useful in the differential diagnosis of hyperthyroidism (see text for details).

and T_3 serum levels that are within the range of normal but are high enough to lead to suppressed serum TSH levels, a situation termed **subclinical hyperthyroidism**. If both FT_4 and TSH are elevated and the RAIU is also elevated, consider a TSH-secreting pituitary tumor or generalized or pituitary-resistance syndromes. Patients who are severely ill may have low serum TSH levels but also have low serum FT_4 and T_3 levels (the so-called *sick euthyroid syndrome*).

In the condition called **familial dysalbuminemic hyperthyroxinemia**, an abnormal albumin is present in serum that preferentially binds T_4 but not T_3. This results in elevation of serum T_4 and FT_4I, and often the free T_4 when measured by usual clinical analogue free T_4 assays, but free T_4 is usually normal by equilibrium dialysis methods, as are serum T_3 and TSH. It is important to differentiate this euthyroid state from true hyperthyroidism. In addition to the absence of clinical features of hyperthyroidism, a normal serum T_3 and a normal TSH level rule out hyperthyroidism.

If eye signs are present, the diagnosis of Graves disease can be made without further tests. If eye signs are absent and the patient is hyperthyroid with or without a goiter, additional testing is needed. An elevated RAIU is typical of Graves disease. In contrast, a low uptake is seen in patients with several forms of spontaneously resolving hyperthyroidism, as in subacute thyroiditis or

silent thyroiditis, also termed **postpartum thyroiditis** (discussed later). A low RAIU is also found in patients who are iodine-loaded or who are taking T_4 therapy—or, rarely, in association with a struma ovarii, where ectopic thyroid tissue is located within an ovarian teratoma. Thyroid autoantibodies—to TG and TPO—are usually present in the serum in both Graves disease and *silent thyroiditis*, a variant of Hashimoto thyroiditis, but TSAb is relatively specific for Graves disease and typically is absent in other forms of thyrotoxicosis. This may be a useful diagnostic test in the patient who presents with unilateral or bilateral exophthalmos without obvious signs or laboratory manifestations of Graves disease (so-called "euthyroid Graves disease"). The ^{123}I or technetium scan is useful to evaluate the size of the gland as well as the presence of hot or cold nodules. CT and MRI scans of the orbits reveal extraocular muscle enlargement in most patients with Graves disease even when there is no clinical evidence of ophthalmopathy. In patients with clinical ophthalmopathy, orbital muscle enlargement may be striking (see Figure 7–36).

Other Presentations

Graves disease occasionally presents in an unusual or atypical fashion, in which case the diagnosis may not be obvious. Marked muscle atrophy may suggest severe myopathy that must be

differentiated from a primary neuromuscular disorder. *Thyrotoxic periodic paralysis* usually occurs in Asian males and presents with a sudden attack of flaccid paralysis and hypokalemia due to a shift of K^+ intracellularly. The paralysis usually subsides spontaneously and can be prevented by K^+ supplementation and β-adrenergic blockade. The illness is cured by appropriate treatment of the thyrotoxicosis (see Chapter 24). Patients with thyrocardiac disease, especially those over age 60, present primarily with symptoms of heart involvement—especially refractory atrial fibrillation insensitive to digoxin—or with high-output heart failure. About 50% of these patients have no evidence of underlying heart disease, and cardiac problems are cured by treatment of the thyrotoxicosis. Some older patients present with weight loss, small goiter, slow atrial fibrillation, and severe depression, with none of the clinical features of increased catecholamine reactivity. These placid patients have *apathetic thyrotoxicosis*. Finally, some young women may present with amenorrhea or infertility as the primary symptom. In all of these instances, the diagnosis of hyperthyroidism can usually be made on the basis of the clinical and laboratory studies described earlier.

Complications

Thyrotoxic crisis (thyroid storm) is the acute exacerbation of all of the symptoms and signs of thyrotoxicosis, often presenting as a syndrome that may be of life-threatening severity (see Chapter 24). Occasionally, thyroid storm may be mild and present simply as an unexplained febrile reaction after thyroid surgery in a patient who has been inadequately prepared. More commonly, it occurs in a more severe form after surgery, radioactive iodine therapy, or parturition in a patient with inadequately controlled thyrotoxicosis—or during a severe, stressful illness or disorder such as uncontrolled diabetes, trauma, acute infection, severe drug reaction, or myocardial infarction. The clinical manifestations of thyroid storm are marked hypermetabolism and excessive adrenergic response. Fever ranges from 38°C to 41°C and is associated with flushing and sweating. There is marked tachycardia, often with atrial fibrillation and high pulse pressure; occasionally heart failure occurs. Central nervous system symptoms include marked agitation, restlessness, delirium, and coma. Gastrointestinal symptoms include nausea, vomiting, diarrhea, and jaundice. A fatal outcome is associated with heart failure and shock.

At one time it was thought that thyroid storm was due to sudden release or dumping of stored T_4 and T_3 from the thyrotoxic gland. Careful studies have revealed, however, that the serum levels of T_4 and T_3 in patients with thyroid storm are not necessarily higher than in thyrotoxic patients without this condition. There is no evidence that thyroid storm is due to excessive production of T_3. In thyrotoxicosis, the number of adrenergic-binding sites for catecholamines increases, so that heart and other tissues have increased sensitivity to circulating catecholamines. In addition, there is decreased binding to TBG, with further elevation of free T_3 and T_4. One theory is that in this setting, with increased binding sites available for catecholamines, an acute illness, infection, or surgical stress triggers an outpouring of catecholamines which, in association with high levels of free T_4 and T_3, precipitates the acute problem.

The most striking clinical diagnostic feature of thyrotoxic crisis is hyperpyrexia out of proportion to other findings. Laboratory findings include elevated serum T_4, FT_4, and T_3 as well as a suppressed TSH (see Chapter 24).

Treatment of Graves Disease

Although autoimmune mechanisms are responsible for the syndrome of Graves disease, management has been largely directed toward controlling the hyperthyroidism. Three good methods are available: (1) antithyroid drug therapy, (2) surgery, and (3) radioactive iodine therapy.

A. Antithyroid drug therapy (see Figure 7–9)   The antithyroid drugs methimazole, carbimazole, and PTU act by inhibiting TPO-mediated iodination of TG to form T_4 and T_3 within the thyroid gland. PTU but not methimazole blocks peripheral T_4 to T_3 conversion; however, this effect is not generally considered to be clinically important except possibly in patients with severe thyrotoxicosis or thyroid storm. Carbimazole, used in the United Kingdom and parts of the British Commonwealth (eg, Australia, South Africa, and Singapore), is a precursor of methimazole that is rapidly metabolized to methimazole. Additionally, each of these drugs may have immunosuppressive effects that may be responsible for the remission from the disease that some patients have after 1 to 2 years of treatment. In general, antithyroid drug therapy as a first-line therapy is most useful in young patients with small glands and mild disease. The drug is given for 1 to 2 years, and then it is tapered or discontinued to see whether the patient has achieved a remission. Remissions, defined as normal thyroid function for 1 year following discontinuation of the antithyroid drug, occur in 20% to 50% of patients but may not be lifelong. Antithyroid drug therapy is generally started with larger doses. When the patient becomes biochemically euthyroid after 4 to 12 weeks, maintenance therapy may be achieved with a lower dose. Because of the potential for severe and potentially fatal hepatotoxicity with PTU, methimazole is the antithyroid drug of choice. Methimazole has fewer adverse reactions, has a longer duration of action, and single daily dosing leads to improved compliance. A typical regimen starts with a 10 to 20-mg dose of methimazole each morning for 1 to 2 months; this dose would then be reduced to 5 to 10 mg each morning for maintenance therapy. PTU could be considered in patients with mild allergic reactions to methimazole who prefer to remain on antithyroid drugs and is preferred in pregnant women in the first trimester because of rare teratogenic effects of methimazole in the fetus. A common regimen consists of giving PTU, 100 mg every 8 hours initially, and then in 4 to 8 weeks, reducing the dose to 50 to 200 mg once or twice daily. The laboratory tests of most value in early monitoring the course of therapy are serum FT_4 and T_3. TSH levels often remain suppressed for many weeks or even months, and hence

are not a reliable index of thyroid function early in the course of treatment.

An alternative method of therapy utilizes the concept of a total block of thyroid activity. The patient is treated with methimazole until euthyroid (about 3-6 months), but instead of continuing to taper the dose of antithyroid drug, at this point T_4 is added in a dose of about 0.1 mg/d. The patient then continues to receive the combination of antithyroid drug and T_4 (0.1 mg/d) for another 12 to 24 months. At the end of this time or when the size of the gland has returned to normal, the drugs are discontinued. This so-called "block-replace" regimen prevents the development of hypothyroidism due to excessive doses of antithyroid drugs, but the frequency of relapse is about the same as after treatment with antithyroid drugs alone. It is more expensive, because two drugs are needed, and there may be a greater frequency of side effects. Consequently, this strategy is not recommended for most patients.

1. **Duration of therapy and other factors related to remissions**—The duration of therapy with antithyroid drugs in Graves disease is usually 12 to 24 months; some patients prefer to be treated, if necessary, for prolonged periods of time (eg, decades) but this is not typical. A sustained remission is most likely in the following circumstances: (1) if the thyroid gland returns to normal size, (2) if the disease can be controlled with a relatively small dose of antithyroid drugs, and (3) if TSAbs are no longer detectable in the serum at the end of the course of therapy.

2. **Reactions to antithyroid drugs**—The adverse reactions to antithyroid drugs almost always occur within the first 3 months of therapy. They can be minor, especially a pruritic rash (about 5% of patients) or major, especially agranulocytosis (about 0.5% of patients). The rash can often be managed by administering antihistamines, and unless it is severe, it is not an indication for discontinuing the medication. Agranulocytosis requires immediate cessation of antithyroid drug therapy, institution of appropriate antibiotic therapy, and shifting to an alternative therapy, usually radioactive iodine. Agranulocytosis is usually heralded by a severe sore throat and fever. Thus, all patients receiving antithyroid drugs are instructed that if sore throat or fever develops, they should stop the drug, contact their physician, and obtain a white blood cell and differential count. If the white blood cell count is normal, the antithyroid drug can be resumed. Cholestatic jaundice with methimazole, hepatocellular toxicity and vasculitis with PTU, and acute arthritis with either drug are serious but rare side effects that also require cessation of drug therapy. As noted earlier, the severity of hepatotoxicity from PTU has led to the warning by the FDA that PTU be used only in special circumstances: in patients who are allergic to methimazole who are not candidates for radioiodine or surgery and in the first trimester of pregnancy. PTU may also be preferred in patients with thyroid storm, where the inhibition of T_4 to T_3 conversion may be clinically important.

B. Radioactive iodine therapy   In the United States, sodium iodide ^{131}I is the preferred treatment for most patients over age 21. In many patients without underlying heart disease, radioactive iodine may be given immediately in a dosage of 80 to 200 μCi/g of thyroid weight estimated on the basis of physical examination. The dosage is corrected for iodine uptake according to the following formula:

$$\text{Dose} = \frac{(80\text{-}200 \text{ microcuries/estimated gram of thyroid tissue}) \times \text{Estimated weight of thyroid (grams)}}{\text{Fractional 24-hour radioiodine uptake}}$$

Following the administration of radioactive iodine, which typically is given orally as a single capsule, the gland will shrink, and the patient will usually become euthyroid over a period of 2 to 6 months.

In elderly patients and in those with underlying heart disease or other medical problems, severe thyrotoxicosis, or large glands (>100 g), it is desirable to achieve a euthyroid state prior to ^{131}I therapy, because there may be occasional exacerbations of thyroid function in the weeks following radioiodine therapy. For this purpose, patients usually are treated with methimazole until they are euthyroid, medication is then stopped for 3 to 7 days, and the radioiodine therapy is given. Methimazole (or PTU) pretreatment have been associated with a diminished effectiveness of radioiodine therapy, which can be overcome by modestly increasing the dose of radioactive iodine.

Hypothyroidism is the almost inevitable complication of radioactive iodine therapy, ultimately developing in more than 80% of patients who are adequately treated. Hypothyroidism usually occurs in the first 6 to 12 months after therapy and is the best assurance that the patient will not have a recurrence of hyperthyroidism. Serum FT_4 and TSH levels should be followed every 6 to 8 weeks, and when hypothyroidism develops, prompt replacement therapy with T_4, 0.05 to 0.2 mg daily, is instituted. Long-term follow-up studies have shown that radioiodine therapy does not cause infertility, birth defects, or cancer later in life. Severe Graves eye disease is a contraindication to radioiodine therapy; several prospective studies have shown that radioiodine can exacerbate eye problems when they are present at baseline, and may even induce mild eye disease in those without eye disease. Patients with severe underlying hyperthyroidism, very high TSAb titers, and who smoke cigarettes are particularly prone to an exacerbation following radioiodine treatment. However, potential worsening can be prevented by the administration of prednisone 0.4 mg/kg/d for 1 to 2 months following the radioiodine treatment. Prednisone is typically used in patients with mild to moderate Graves eye disease, especially in those who smoke cigarettes.

Hypothyroidism may also occur as a late development after an earlier course of antithyroid drug therapy for Graves disease; in such patients, *burned-out* Graves disease is likely the end result of autoimmune thyroid destruction. Accordingly, all patients with Graves disease who have not become hypothyroid require lifetime follow-up to be certain that they remain euthyroid.

C. Surgical treatment   Total or near-total thyroidectomy is the treatment of choice for patients with very large glands or multinodular goiters, for patients with a concomitant suspicious or malignant thyroid nodule, for patients who are allergic to or

noncompliant with antithyroid drugs, for patients who refuse radioactive iodine, and for pregnant women with severe Graves disease who are allergic or develop reactions to antithyroid drugs. The patient is prepared with antithyroid drugs until euthyroid (about 6 weeks). In addition, starting 2 weeks before the day of operation, the patient may be given saturated solution of potassium iodide, 5 drops twice daily. This regimen is thought to diminish the vascularity of the gland and thereby decrease intraoperative blood loss.

An experienced surgeon is required, and total thyroidectomy is recommended, because if too much thyroid tissue is left behind, the disease may recur. Hypoparathyroidism and recurrent laryngeal nerve injury occur as complications of surgery in about 1% of cases, with lower rates in the hands of *high volume* surgeons.

D. Other medical measures During the acute phase of thyrotoxicosis, β-adrenergic–blocking agents are extremely helpful. Propranolol, 10 to 40 mg every 6 hours, or longer acting beta-blockers such as Inderal LA, nadolol, atenolol, or metoprolol control many adrenergic symptoms, tachycardia, hypertension, and atrial fibrillation. The drug is gradually withdrawn as serum FT_4 and T_3 levels return to normal. Adequate nutrition and rest are both essential. The oral cholecystographic dyes, sodium ipodate and iopanoic acid (no longer available in the United States), have been shown to inhibit both thyroid hormone synthesis and release as well as peripheral conversion of T_4 to T_3. In a dosage of 1 g daily, these drugs may help to rapidly restore the euthyroid state in very ill patients. They leave the gland saturated with iodide, so they should not be used immediately before ^{131}I therapy. Cholestyramine, 4 g orally 3 times daily, lowers serum T_4 by binding it in the gut, and may be useful in severely hyperthyroid patients.

Choice of Therapy

The choice of therapy varies with the nature and severity of the illness and prevailing customs. For example, in the United States, radioiodine therapy has been the preferred treatment for the average patient, whereas in Europe, Latin America, and Asia, primary antithyroid drug therapy is preferred. In general, antithyroid drugs are a reasonable initial therapy for children and adolescents and in adults with mild disease and small goiters. In all other patients, radioiodine should be strongly considered. In contemporary U.S. clinical practice, the role of surgery is limited to noncompliant patients who refuse radioiodine, patients with large goiters, and women desirous of pregnancy in the near future who do not want to take antithyroid drugs in pregnancy.

Treatment of Complications

A. Thyrotoxic crisis Thyrotoxic crisis (thyroid storm) requires vigorous management. Propranolol, 1 to 2 mg slowly intravenously every 5 to 10 minutes for a total of 10 mg, or 40 to 80 mg every 6 hours orally, is helpful in controlling arrhythmias. Large doses of propranolol also have the ability to block T_4 to T_3 conversion. Esmolol, a very rapidly acting beta-blocker, is given at a dose of 250 to 500 mg/kg/min as a loading dose followed by 50 μg/kg/min with cardiac monitoring. In the presence of severe heart failure or asthma and arrhythmia, where beta-blockers may be contraindicated, cautious intravenous administration of verapamil in a dose of 5 to 10 mg may be effective. Hormone synthesis is blocked by the administration of PTU, 250 mg every 6 hours. In thyroid storm, PTU may be preferable to methimazole because PTU blocks T_4 to T_3 conversion. If the patient is unable to take medication by mouth, methimazole in a dose of 60 mg every 24 hours or PTU, 400 mg every 6 hours, can be given by rectal suppository or enema.[4] After administration of an antithyroid drug, hormone release is retarded by the oral administration of a saturated solution of potassium iodide, 10 drops twice daily. As noted earlier, the oral cholecystographic agents, sodium ipodate (Oragrafin) or iopanoic acid (Telepaque), provide a source of iodine and potently block T_4 to T_3 conversion, but these drugs are not available currently in the United States. The administration of hydrocortisone hemisuccinate, or its equivalent, 50 mg intravenously every 6 hours, is usually recommended, but its role in thyroid storm management is uncertain. Supportive therapy includes a cooling blanket and acetaminophen to help control fever. Aspirin is contraindicated because of its tendency to bind to TBG and displace T_4, rendering more T_4 available in the free state. Proper fluid, electrolyte, and nutritional support are important. For sedation, phenobarbital may be useful because it accelerates the peripheral metabolism and inactivation of T_4 and T_3. Oxygen, diuretics, and digoxin are indicated for heart failure and/or atrial fibrillation. Finally, it is essential to treat the underlying disease process that may have precipitated the acute exacerbation. As an extreme measure (rarely needed) to control thyrotoxic crisis, plasmapheresis or peritoneal dialysis may be used to remove high levels of circulating-free thyroid hormones (see Chapter 24).

B. Ophthalmopathy For mild disease, keeping the patient's head elevated at night and the administration of diuretics may help diminish periorbital edema. As noted earlier, prednisone begun immediately after radioiodine in a dose of 0.4 mg/kg/d, reducing the dose by 10 mg every 2 weeks, protects against exacerbation of ophthalmopathy following ^{131}I therapy. Management of severe ophthalmopathy due to Graves disease involves close cooperation between the endocrinologist and the ophthalmologist. For severe acute inflammatory reactions, a

---

[4] **Preparation of rectal methimazole:** Dissolve 1200 mg methimazole in 12 mL of water to which a mixture of two drops of Span 80 in 52 mL of cocoa butter warmed to 37°C has been added. Stir the mixture to form a water-oil emulsion, pour into 2.6 mL suppository molds, and cool. Each suppository supplies approximately 60 mg methimazole-absorbed dose (Nabil et al., 1982).
Preparation of rectal PTU: Dissolve 400 mg PTU in 60 mL of Fleet Mineral Oil for the first dose and then dissolve 400 mg PTU in 60 mL of Fleet Phospho-Soda for subsequent enemas.

short course of high-dose corticosteroid therapy is frequently effective. Weekly pulse therapy with high doses of intravenous methylprednisolone (500 mg weekly for 6 weeks followed by 250 mg weekly for 6 weeks) has been shown to be more effective and less toxic than oral prednisone therapy. If corticosteroid therapy is not effective or if there is recurrence after the drug is tapered, external x-ray therapy to the retrobulbar area may be helpful. The dose is usually 2000 cGy in 10 fractions given over a period of 2 weeks. The lens and anterior chamber structures must be shielded.

In very severe cases or where vision is threatened, orbital decompression can be used. One operation involves a transantral approach through the maxillary sinus, removing the floor and the lateral walls of the orbit. In the alternative anterior approach, the orbit is entered under the globe, and portions of the floor and the walls of the orbit are removed. Both approaches are effective, and exophthalmos can be reduced by 5 to 7 mm in each eye by these techniques. After the acute process has subsided, the patient is frequently left with double vision or lid abnormalities owing to muscle fibrosis and contracture or the postoperative position of the globe. These problems can be corrected by cosmetic lid surgery or eye muscle surgery.

C. Thyrotoxicosis and pregnancy   Thyrotoxicosis during pregnancy presents a special problem. It is unusual, affecting approximately 0.1% of pregnancies. It is important to recall that serum TSH levels may be subnormal at the end of the first trimester in up to 20% of normal women. This is due to the effects of serum hCG, the levels of which normally peak at the end of the first trimester; hCG activates the TSH receptor. This is a transient phenomenon and is physiologic, not pathologic. Hyperemesis gravidarum may cause mild hyperthyroidism (so-called *gestational thyrotoxicosis*), which is likely due to extremely high-serum levels of hCG, present in this condition, that stimulates the TSH receptor. However, no treatment is indicated, because the hyperthyroid state resolves when the hyperemesis resolves, usually by midgestation.

In pregnant patients with Graves disease, the patient is treated with antithyroid drugs throughout most of the pregnancy, postponing the decision regarding long-term management until after delivery. Radioactive iodine is absolutely contraindicated because it crosses the placenta freely and may injure the fetal thyroid. The dosage of antithyroid drugs must be kept to the minimum necessary to control symptoms, because both PTU and methimazole cross the placenta and may affect the function of the fetal thyroid gland. PTU is preferred over methimazole because methimazole has rare severe teratogenic effects (aplasia cutis and methimazole embryopathy [choanal atresia, tracheoesophageal fistulae, and other defects]). While PTU use has also been associated with birth defects, these are not as severe (eg, renal cysts and preauricular sinuses). Current recommendations advise limiting PTU use to the first trimester, and then switching to methimazole. However, methimazole can be used when PTU allergy or a poor clinical response to PTU is present. If the disease can be controlled by initial doses of PTU of 250 mg/d (in divided doses) or less and

maintenance doses of 25 to 100 mg/d, the likelihood of fetal hypothyroidism is extremely small. The FT_4I or FT_4 should be maintained in the upper range of normal for nonpregnant women by appropriately reducing the antithyroid drug dosage. In a significant minority of patients, the drug can be discontinued in the latter part of pregnancy because of spontaneous disappearance of TSAb. Serum levels of TSAb should be measured in the third trimester; high titers can be associated with neonatal Graves disease (see later). Breastfeeding is permissible with either antithyroid drug, because the low levels found in breast milk do not affect thyroid function in the neonate.

Graves disease may occur in the newborn infant (*neonatal Graves disease*) due to transplacental passage of TSAb. Although most infants with this rare syndrome are born to mothers with active Graves disease, it occasionally can occur in infants born to hypothyroid mothers who have been treated for Graves disease with radioiodine in the past but who still retain circulating TSAb. The disease may not be evident at birth, because the antithyroid drugs used to treat the mother can cross the placenta and affect the infant's thyroid as well. The child is born small, with weak muscles, tachycardia, fever, and frequently with distress or neonatal jaundice. Examination reveals an enlarged thyroid gland and occasionally prominent, puffy eyes. The heart rate is rapid, temperature is elevated, and heart failure may ensue. Laboratory studies reveal an elevated FT_4I or FT_4, a markedly elevated T_3, and usually a low TSH—in contrast to normal infants, who have elevated TSH at birth. Bone age may be accelerated. TSAb is usually found in the serum of both the infant and the mother. The pathogenesis of this syndrome is thought to involve transplacental transfer of TSAb from mother to fetus, with subsequent development of thyrotoxicosis. The disease is self-limited and subsides over a period of 4 to 12 weeks, coinciding with the fall in the child's serum TSAb levels. Therapy for the infant includes PTU in a dose of 5 to 10 mg/kg/d (in divided doses at 8-hour intervals); Lugol solution, one drop (8 mg potassium iodide) every 8 hours; and propranolol, 2 mg/kg/d in divided doses. In addition, adequate nutrition, antibiotics for infection if present, sedatives if necessary, and supportive therapy are indicated. If the child is very thyrotoxic, corticosteroid therapy (prednisone, 2 mg/kg/d) partially blocks conversion of T_4 to T_3 and may be helpful in the acute phase. The above medications are gradually reduced as the child improves and can usually be discontinued by 6 to 12 weeks.

Maternal sera may also contain TSAb that functions as blocking Abs that can cross the placenta to produce transient hypothyroidism in the infant. This condition may need to be treated with T_4 supplementation for a short time.

Course and Prognosis

In general, the course of Graves disease treated with antithyroid drugs is one of remissions and exacerbations over a protracted period of time, unless the gland is destroyed by surgery or radioactive iodine. Although some patients may remain euthyroid for long periods after antithyroid drug treatment, at least 25% eventually develop hypothyroidism. Lifetime follow-up is therefore indicated for all patients with Graves disease.

2. OTHER FORMS OF THYROTOXICOSIS

Toxic Adenoma

A functioning adenoma hypersecreting T_3 and T_4 causes hyperthyroidism. These lesions start out as a small autonomously functioning nodule that slowly increases in size to produce excessive quantities of thyroid hormones. This gradually suppresses endogenous TSH secretion, which results in reduced function of the contralateral lobe of the gland (Figure 7–39). The typical patient is an older individual (usually >40) who has noted recent growth of a long-standing thyroid nodule. Symptoms of weight loss, weakness, shortness of breath, palpitations, tachycardia, and heat intolerance are noted. Infiltrative ophthalmopathy is never present.

Physical examination reveals a definite nodule on one side, with very little thyroid tissue on the other side. Laboratory studies usually reveal suppressed TSH and elevation in serum T_3 levels, often with only borderline elevation of FT_4 levels. A radionuclide scan reveals that the nodule is hot, with diminished or absent function of the contralateral lobe. Toxic adenomas are usually benign follicular adenomas and almost never malignant.

Treatment with radioactive iodine is generally effective and is attractive because the contralateral lobe can be spared the destructive effects of the radiation, and the patient should therefore remain euthyroid. Radioiodine therapy in doses of 20 to 30 mCi is usually required to destroy the benign neoplasm. If antithyroid drugs are used to pretreat the patient, it is important that the serum TSH remain low, so that the contralateral lobe will not be stimulated to take up the radioiodine. Surgery is an option, if the nodule is very large and causing obstructive symptoms such as dysphagia, neck pressure, or difficulty breathing. Antithyroid drugs may be used to normalize thyroid function prior to radioiodine or surgery but are not a good long-term solution, because, unlike the situation with Graves disease, remissions do not occur.
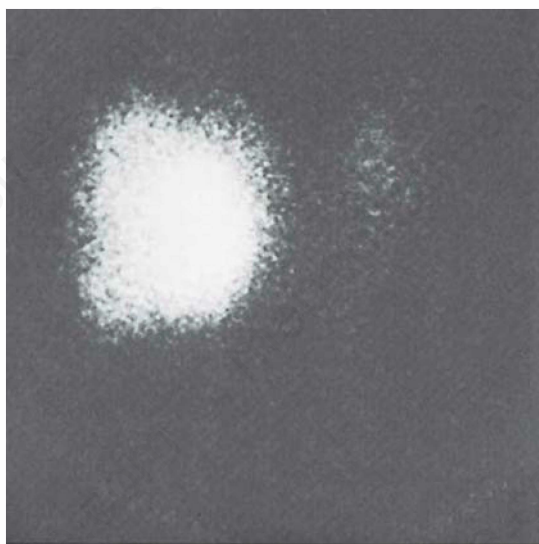
Toxic Multinodular Goiter (Plummer Disease)

This disorder usually occurs in older patients with longstanding euthyroid multinodular goiter. The patient presents with tachycardia, heart failure, or arrhythmia and sometimes weight loss, nervousness, weakness, tremors, and sweats. Physical examination reveals a multinodular goiter that may be small or quite large and may even extend substernally. Laboratory studies reveal a suppressed TSH and elevation in serum T_3 levels, with less striking elevation of serum T_4. Radioiodine scan reveals multiple functioning nodules in the gland or occasionally an irregular, patchy distribution of radioactive iodine (Figure 7–40).

Hyperthyroidism in patients with longstanding multinodular goiters can be precipitated by the administration of pharmacologic doses of iodide-containing drugs (called the Jod-Basedow phenomenon or iodide-induced hyperthyroidism). The pathophysiology of iodine-induced hyperthyroidism is unknown but likely involves the inability of some thyroid nodules to adapt to an iodide load, with resulting excess hormone production. This is one mechanism for the development of hyperthyroidism after administration of the iodine-containing antiarrhythmic drug amiodarone (see later).

The management of toxic nodular goiter may be difficult, because patients are often elderly with other comorbidities. Control of the hyperthyroid state with antithyroid drugs followed by radioiodine is the therapy of choice. If the goiter is very large, thyroidectomy can be contemplated if the patient is a good surgical candidate.

Amiodarone-Induced Thyrotoxicosis

Amiodarone is an antiarrhythmic drug that contains 37.3% iodine by weight. In the body, it is stored in fat, myocardium,

FIGURE 7–39 Solitary toxic nodule as it appears on ^{99m}Tc pertechnetate scan. Note suppression of contralateral lobe (left) by toxic nodule (right).

FIGURE 7–40 Toxic multinodular goiter as it appears on ^{99m}Tc pertechnetate scan. Note multiple functioning thyroid nodules. (Used with permission from Dr. J.M. Lowenstein.)

liver, and lung and has a half-life of about 50 days. In the United States, about 2% of patients treated with amiodarone develop amiodarone-induced thyrotoxicosis. In some patients, the thyrotoxicosis is due to the excess iodine (also called type 1 amiodarone-induced thyrotoxicosis); in others, it is due to an amiodarone-induced thyroiditis, with inflammation and release of stored hormone into the bloodstream as the result of thyroidal inflammation (also termed type 2 amiodarone-induced thyrotoxicosis). In general, the thyrotoxicosis associated with excess iodine occurs within 6 to 12 months of starting amiodarone therapy, while patients developing thyroiditis have usually been taking the drug for 2 to 3 years. Thyroid ultrasound with Doppler examination of the thyroid circulation may be helpful in differentiating the two syndromes: the thyroidal blood flow is increased in iodide-induced hyperthyroidism and diminished in thyroiditis. Treatment is often difficult. Iodide-induced thyrotoxicosis can be controlled with methimazole, 40 to 60 mg/d, and β-adrenergic blockade. Although potassium perchlorate (250 mg every 6 hours) could be added to block further iodide uptake, long-term, high-dose treatment with this agent has been associated with aplastic anemia and requires careful monitoring. Also, this drug is no longer available in the United States. The drug-induced thyroiditis responds to prednisone therapy, which may need to be continued for several months. Some patients have what appears to be a mixed form of disease and respond to a combination of antithyroid drugs and prednisone. Total thyroidectomy is curative and may be needed in patients who are nonresponsive to pharmacologic therapy. Of course, surgery is feasible only if the patient can withstand the stress of surgery (see Chapter 24).

Subacute and Silent Thyroiditis

These entities will be discussed in a separate section, but it should be mentioned here that thyroiditis, either subacute or silent, may present with symptoms of mild to severe thyrotoxicosis following an acute release of T_4 and T_3 into the circulation. These illnesses can be differentiated from other forms of thyrotoxicosis in that the RAIU is markedly suppressed, and the symptoms usually subside spontaneously over a period of weeks or months.

Thyrotoxicosis Factitia

This is a psychoneurotic disturbance in which the patient surreptitiously ingests excessive amounts of T_4 or thyroid hormone preparations, usually for purposes of weight control. The individual, usually a woman, is often someone connected with the health-care field who can easily obtain thyroid medication. Features of thyrotoxicosis, including weight loss, nervousness, palpitation, tachycardia, and tremor, may be present, but goiter and eye signs are absent. Characteristically, TSH is suppressed, serum FT_4 and T_3 levels are elevated, serum TG is low, and RAIU is nil. Management requires careful discussion of the hazards of long-term thyroid hormone therapy, particularly cardiovascular damage, muscle wasting, and osteoporosis. Formal psychotherapy may be necessary.

Rare Forms of Thyrotoxicosis

A. Struma ovarii   In this syndrome, a teratoma of the ovary contains thyroid tissue that becomes hyperactive. Mild features of thyrotoxicosis result, such as weight loss and tachycardia, but there is no evidence of goiter or eye signs. Serum FT_4 and T_3 are mildly elevated, serum TSH is suppressed, and RAIU over the neck is low. Total body scan reveals uptake of radioiodine in the pelvis, rather than in the neck. The disease is curable by removal of the teratoma.

B. Thyroid carcinoma   Follicular carcinoma of the thyroid may concentrate radioactive iodine, but only rarely does it retain the ability to convert this iodide into active hormone. Only a few cases of metastatic thyroid cancer are described in which patients have presented with hyperthyroidism. The clinical picture consists of weakness, weight loss, palpitations, and a thyroid nodule but no ophthalmopathy. Body scan with ^{131}I reveals areas of uptake usually distant from the thyroid (eg, bone or lung). Treatment with large doses of radioactive iodine may destroy the metastatic deposits and treat the hyperthyroidism.

C. Hydatidiform mole and choriocarcinoma   Hydatidiform moles and choriocarcinomas produce high levels of hCG, which has intrinsic TSH-like activity. This may induce thyroid hyperplasia, increased RAIU, suppressed TSH, and mild elevation of serum T_4 and T_3 levels. It is rarely associated with overt thyrotoxicosis and is curable by removal of the mole or treatment of the tumor.

D. Hamburger thyrotoxicosis   In the 1970s an epidemic of thyrotoxicosis in the midwestern United States was traced to hamburger made from *neck trim*, the strap muscles from the necks of slaughtered cattle that contained beef thyroid tissue. The United States Department of Agriculture has now prohibited the use of this material for human consumption.

E. Syndrome of inappropriate TSH secretion   A group of patients have been reported with elevated serum FT_4 concentrations in association with elevated or inappropriately normal serum immunoreactive TSH levels. This has been called the **syndrome of inappropriate TSH secretion**. Two conditions may explain this unusual combination of laboratory values: (1) TSH-secreting pituitary adenoma, and (2) non-neoplastic pituitary hypersecretion of TSH.

Patients with TSH-secreting pituitary adenomas usually present with mild thyrotoxicosis and goiter and may have a concurrent pituitary hormonal deficiency, such as central hypogonadism causing amenorrhea or impotence. Rarely, a TSH-secreting tumor may cosecrete other pituitary hormones such as growth hormone or prolactin. There are no eye signs of Graves disease. Laboratory evaluation reveals elevated FT_4 and T_3, whereas serum TSH, usually undetectable in Graves disease, is within the normal range or even elevated. The TSH α subunit secretion from these tumors is markedly elevated; a molar ratio of α subunit: TSH greater than

5.7 is usually diagnostic of the presence of a TSH-secreting pituitary adenoma.[5]

In addition, the increased serum TSH is not suppressible with high doses of exogenous thyroid hormones. Visual field examination may reveal a bitemporal defect, and CT or MRI of the sella usually reveals a pituitary tumor. Management usually involves control of the thyrotoxicosis with antithyroid drugs and removal of the pituitary tumor via transsphenoidal hypophysectomy. These tumors are often quite aggressive and may extend widely out of the sella. If the tumor cannot be completely removed, it may be necessary to treat residual tumor with radiation therapy and to control thyrotoxicosis with radioactive iodine. Long-acting somatostatin analog (octreotide) suppresses TSH secretion in many of these patients and may even inhibit tumor growth in some.

Non-neoplastic pituitary hypersecretion of TSH is essentially a form of tissue resistance to T_3 and T_4. This is discussed in the following section.

Resistance to Thyroid Hormone Syndromes

Thyroid hormone action is mediated by nuclear thyroid hormone receptors (TRs), TRα and TRβ, which are encoded by the *THRA* and *THRB* genes, respectively. Two forms of RTHs are recognized. The first is GRTH, which can be caused by a number of different mutations in *THRB*, which confers resistance to TRβ, leading to elevated serum levels of the thyroid hormones and a TSH level that is inappropriately normal or slightly elevated. The second is resistance to thyroid hormone receptor alpha (RTHα) caused by *THRA* mutations, which are associated with a low serum TSH and slightly low FT_4 level. In both conditions, most patients are heterozygous for a single mutant TR gene (either THRB or THRA) that produces a mutant protein capable of inhibiting thyroid hormone actions by a dominant negative mechanism (see later).

GRTH was first described in 1967 as a familial syndrome of deaf mutism, stippled epiphyses, goiter, and high circulating thyroid hormone levels with normal TSH. The clinical presentation as reported in more than 500 cases has been variable. Although most affected individuals are clinically euthyroid; some present with goiter, stunted growth, delayed maturation, attention deficits, hyperactivity disorders, and resting tachycardia. The disorder is familial in 75% of cases, and inheritance is autosomal dominant. Laboratory tests reveal elevated serum T_4, FT_4, and T_3 with normal or elevated TSH. Testing to distinguish GRTH from TSH-secreting adenomas usually reveals a fall in TSH with T_3 administration, a molar ratio of α subunit-to-TSH of less than 1 and a pituitary MRI without evidence of a pituitary tumor in GRTH. Molecular studies have revealed point mutations in the carboxyl terminal ligand-binding portion of TRβ, which produces a defective TR that demonstrates reduced binding to T_3 while retaining the ability to bind to DNA. The mutant TR exerts a dominant negative action by occupying the TRE as either an inactive homodimer or heterodimer, both of which repress gene expression (see Figures 7–20 and 7–21). While distinct point mutations in different families may account, in part, for differences in clinical expression of the syndrome, there can be considerable phenotypic variation in a single kindred. Identifying the specific mutation in affected individuals does allow use of molecular screening methods for the diagnosis of the syndrome in other family members.

In most patients with GRTH, the increased levels of T_3 and T_4 appear to compensate largely for the receptor defect, and treatment is unnecessary. In fact, some patients have been erroneously thought to have Graves disease and inappropriately undergone therapy with radioiodine or surgery. However, in some children, administration of thyroid hormones may be necessary to correct defects in growth or mental development.

Some individuals with GRTH have symptoms of mild hyperthyroidism suggesting that their degree of thyroid hormone resistance varies from tissue to tissue. This is consistent with the observation that identical *THRB* gene mutations have been found in patients with various degrees of clinical thyroid hormone resistance. If the hyperthyroid state in such individuals is causing significant symptoms, therapy with beta-adrenergic–blocking agents, antithyroid drugs, or even radioactive iodine can be considered.

The first patient with RTHα was reported in 2012. This individual, and subsequently several others, have delayed growth, failed tooth eruption, constipation, low intellectual function, and decreased heart rate. These clinical features are consistent with an abnormal TRα protein, which is expressed predominantly in the skeleton, heart, and likely the intestine. Affected individuals' thyroid function tests typically include a low-normal FT_4, high-normal or elevated T_3, and normal TSH.

TSH Receptor Gene Mutations

Mutations in the *TSH receptor* (*TSH-R*) gene can produce a variety of clinical syndromes. Somatic mutations in the seven-transmembrane loop of the *TSH-R* may activate the receptor, producing solitary or multiple hyperfunctioning adenomas, whereas germline mutations may result in congenital hyperthyroidism in the newborn that can mimic neonatal Graves disease. Mutations in the extracellular amino terminal portion of the *TSH-R* produce resistance to TSH with uncompensated or compensated hypothyroidism. Some patients may have normal serum FT_4 and T_3 levels and normal growth and development but persistently elevated serum TSH. Others may be severely hypothyroid (cretinoid), with low FT_4 levels, elevated TSH, and no response to exogenous TSH. In this group, the defect may be in the coupling of *TSH-R* and the G_s protein necessary for activation of adenylyl cyclase.

NONTOXIC GOITER

Etiology

Nontoxic goiter (ie, goiter not associated with hyperthyroidism) can be diffuse or nodular. In some circumstances, it results from

---

[5]The α subunit-TSH molar ratio is calculated as follows: (α subunit in μg/L divided by TSH in mU/L) × 10. Normal range (for a patient with normal TSH and gonadotrophins) is less than 1.

TSH stimulation, which in turn results from inadequate thyroid hormone synthesis. Some goiters are due to mutations in genes involved in thyroid growth and/or thyroid function. In many patients, however, the cause of the goiter is obscure, because serum TSH levels are normal. Worldwide, iodine deficiency remains the most common cause of nontoxic goiter or *endemic goiter*. With the widespread use of iodized salt and the introduction of iodides into fertilizers, animal feeds, and food preservatives, iodide deficiency in developed countries has become relatively rare, and it probably does not exist in the United States. However, there are large areas of Africa, the mountainous areas of central Asia, parts of central Europe, and Papua New Guinea, where iodine intake is still deficient. Even highly developed countries, including Australia and the United Kingdom, have recently been found to have mild to moderate iodine deficiency. Optimal iodine requirements for adults are in the range of 150 to 300 µg/d. In endemic goiter areas, the daily intake (and urinary excretion) of iodine falls below 50 µg/d; in areas where iodine is extremely scarce, excretion falls below 20 µg/d. It is in these areas that 90% of the population has goiters, and 5% to 15% of infants are born with myxedematous or neurologic changes of cretinism. The variability in the extent of goiter in these areas may be related to the presence of other, unidentified goitrogens such as goitrin, an organic compound found in certain roots and seeds; and cyanogenic glycosides, found in cassava and cabbage, that release thiocyanate that can exacerbate the effects of iodide deficiency. In addition, compounds such as phenols, phthalates, pyridines, and polyaromatic hydrocarbons found in industrial waste water are weakly goitrogenic.

The most common cause of thyroid enlargement in developed countries is chronic thyroiditis (Hashimoto thyroiditis; see later). The cause of goiter not due to autoimmune thyroid disease or iodine sufficiency is uncertain. In some patients, mild defects in thyroid hormone synthesis (dyshormonogenesis) (discussed earlier in the section Hypothyroidism) may result in goiter with relatively normal thyroid function. Finally, thyroid enlargement can be due to gene mutations producing a benign lesion, such as adenoma, or to a malignant one, such as carcinoma.

Pathogenesis

The development of nontoxic goiter in patients with dyshormonogenesis or severe iodine deficiency involves impaired hormone synthesis and, secondarily, an increase in TSH secretion. TSH induces diffuse thyroid hyperplasia, followed by focal hyperplasia with necrosis and hemorrhage, and finally the development of new areas of focal hyperplasia. Focal or nodular hyperplasia usually involves a clone of cells that may or may not be able to concentrate iodine or synthesize TG. Thus, the nodules vary from hot nodules that can concentrate iodine to cold ones that cannot, and from colloid nodules that can synthesize TG to microfollicular ones that cannot. Initially, the hyperplasia is TSH dependent, but later the nodules become TSH independent, or autonomous. Thus, a diffuse nontoxic TSH-dependent goiter may progress over an extended period of time to become a toxic multinodular TSH-independent goiter.

The mechanism for the development of autonomous growth and function of thyroid nodules may involve mutations that activate the G_s protein in the cell membrane. Mutations of this gene, called the *gsp* oncogene, have been found in a high proportion of nodules from patients with multinodular goiter. Chronic activation of the G_s protein results in thyroid cell proliferation and hyperfunction even when TSH is suppressed.

In iodine-sufficient parts of the world such as the United States, euthyroid nontoxic goiter is a very common problem affecting up to 15% of women. The cause of goiter in the absence of iodine deficiency, autoimmunity, or obvious biosynthetic defects is unknown. In some kindreds with euthyroid multinodular goiter, mutations in the *TG* gene have been identified, suggesting the presence of subclinical or mild defects in thyroid hormone synthesis that do not cause overt hypothyroidism or even elevation in serum TSH levels.

Clinical Features

A. Symptoms and signs   Patients with nontoxic goiter usually present with thyroid enlargement, which, as noted earlier, may be diffuse or multinodular. The gland may be relatively firm but is often soft or rubbery in consistency. Over a period of time, the gland may become progressively larger, so that in long-standing cases, huge goiters may develop and extend substernally to the level of the aortic arch. Facial flushing and dilation of cervical veins on lifting the arms over the head is a positive **Pemberton sign** and indicates obstruction to jugular venous flow (Figure 7–41). The patient may complain of pressure symptoms in the neck, particularly on moving the head upward or downward, and of difficulty in swallowing. Vocal cord paralysis due to recurrent laryngeal nerve involvement is rare. The vast majority of patients are euthyroid. Thyroid enlargement probably represents compensated hypothyroidism.

B. Laboratory findings   Laboratory studies reveal a normal FT_4 and, usually, normal levels of TSH. The increased mass of thyroid tissue may compensate for inefficient synthesis of hormone. RAIU may be high, normal, or low, depending on the iodide pool and the serum TSH level.

C. Imaging studies   Isotope scanning usually reveals a patchy uptake, frequently with focal areas of increased uptake corresponding to hot nodules and areas of decreased uptake corresponding to cold nodules. Thyroid ultrasound is a simple way to follow the growth of the goiter and in addition may reveal cystic changes and coarse calcifications in one or more of the nodules, representing previous hemorrhage and necrosis. Cross sectional imaging with CT scan allows the delineation of tracheal deviation or compression as well as possible substernal extension of the goiter.

Differential Diagnosis

The major problem in differential diagnosis is to rule out cancer. This will be discussed in the section Thyroid Carcinoma.

FIGURE 7–41   Pemberton sign (facial plethora) when the arms are raised above the head due to obstruction of jugular venous flow as the thyroid is forced downward into the sternal notch. This patient has a large goiter shown on an MRI. (Reproduced with permission from Basaria S, Salvatori R. Pemberton sign. *N Engl J Med*. 2004 Mar 25;350(13):1338.)

Treatment

With the exception of those due to neoplasm, the current management of nontoxic goiters consists simply of observation, without any specific therapy. Nontoxic goiters are very slow growing and usually do not cause compressive symptoms or thyroid dysfunction. Thyroid hormone suppression therapy rarely results in clinically significant decrements in goiter size. Also, longstanding goiters may have areas of necrosis, hemorrhage, and scarring as well as autonomously functioning nodules that will not regress on T_4 therapy. Furthermore, to suppress the serum TSH level, doses of T_4 are required that could cause harm, especially in elderly patients who are at risk of developing atrial fibrillation and osteoporosis. Also, many nontoxic goiters have areas of autonomy within them that do not depend on TSH for function, and therefore, will not decrease in size and may contribute to the iatrogenic hyperthyroidism.

Surgery is indicated for goiters that continue to grow or that produce obstructive symptoms. Substernal extension of a goiter is not, in and of itself, an indication for surgical removal. A large multinodular goiter with tracheal deviation and substernal extension is shown in Figure 7-42. Although multinodular goiters are rarely malignant, the size of the mass with resulting pressure symptoms may require subtotal thyroidectomy.

FIGURE 7–42 Massive multinodular goiter (arrow) with substernal extension causing tracheal narrowing and deviation to the right.

If the patient is not a suitable candidate for surgery, radioiodine ablation of functioning thyroid tissue may provide palliative relief of obstructive symptoms. An adequate dose of radioiodine reduces the size of the goiter about 30% to 50% and usually alleviates obstructive symptoms.

Course and Prognosis

Patients with nontoxic goiter should avoid pharmacologic doses of iodides, which may induce either hyperthyroidism or, in the presence of thyroidal autoimmunity, hypothyroidism. Occasionally, single adenomas or several adenomas become hyperfunctional, producing a toxic nodular goiter (discussed earlier). Nontoxic goiter is often familial and other members of the family should be examined and observed for the possible development of goiter.

THYROIDITIS

1. SUBACUTE THYROIDITIS

Subacute thyroiditis (De Quervain thyroiditis or granulomatous thyroiditis) is an acute inflammatory disorder of the thyroid gland most likely due to viral infection. A number of viruses, including mumps virus, coxsackievirus, and adenoviruses, have been implicated, either by finding the virus in biopsy specimens taken from the gland or by demonstration of rising titers of viral Abs in the blood during the course of the infection. Pathologic examination reveals moderate thyroid enlargement and a mild inflammatory reaction involving the capsule. Histologic features include destruction of thyroid parenchyma and the presence of many large phagocytic cells, including giant cells. Subacute thyroiditis is more common in the summer months, in women, and in individuals who are HLA-Bw35 positive.

Clinical Features

A. Symptoms and signs   Subacute thyroiditis usually presents with fever, malaise, and soreness in the anterior neck, which may extend up to the angle of the jaw or toward the ear lobes on one or both sides of the neck. Initially, the patient may have symptoms of hyperthyroidism, with palpitations, nervousness, and sweats. There is no ophthalmopathy. On physical examination, the gland is exquisitely tender, so that the patient objects to pressure on it. There are no signs of local redness or heat suggestive of abscess formation. Clinical signs of thyrotoxicosis, including tachycardia, tremor, and hyperreflexia, may be present.

B. Laboratory findings   Laboratory studies vary with the course of the disease (Figure 7–43). Initially, FT_4 and T_3 are elevated, whereas serum TSH and RAIU are extremely low. Because the thyroid hormone levels in the blood are a reflection of leakage of glandular hormonal stores, serum FT_4 levels are elevated disproportionately to serum T_3 levels. The erythrocyte sedimentation rate is markedly elevated, sometimes greater than 100 mm/h by the Westergren scale. Thyroid autoantibodies are usually not detectable in serum. As the disease progresses, FT_4 and T_3 drop, TSH rises, and symptoms of hypothyroidism may be noted. Later, RAIU rises, reflecting recovery of the gland from the acute insult.

Differential Diagnosis

Subacute thyroiditis can be differentiated from other viral illnesses by the involvement of the thyroid gland. It is differentiated from
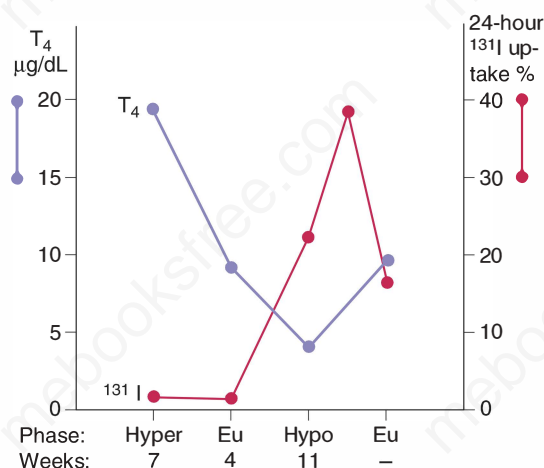

FIGURE 7–43 Changes in serum T_4 and radioactive iodine uptake in patients with subacute, silent, and postpartum thyroiditis. In the initial phase, serum T_4 is elevated and the patient may have symptoms of thyrotoxicosis, but radioactive iodine uptake is markedly suppressed. The illness may pass through phases of euthyroidism and hypothyroidism before remission. (Data from Woolf PD, Daly R. Thyrotoxicosis with painless thyroiditis. *Am J Med*. 1976 Jan; 60(1):73-79.)

Graves disease by the presence of thyroidal pain, low RAIU associated with elevated serum T_3 and FT_4 and suppressed serum TSH, and by the absence of thyroid Abs.

Treatment

In many cases, only symptomatic treatment is necessary (eg, aspirin or other nonsteroidal anti-inflammatory drugs). In severe cases or in patients who do not respond to nonsteroidal agents, a glucocorticoid such as prednisone, 20 mg three times daily for 7 to 10 days, may be necessary to reduce the inflammation. β-Adrenergic–blocking agents can be employed during the hyperthyroid phase to treat bothersome symptoms of hyperthyroidism.

T_4, 0.1 to 0.15 mg once daily, may be indicated during the hypothyroid phase of the illness if hypothyroid symptoms are present. T_4 therapy may also prevent exacerbations of inflammation due to elevated TSH levels.

Course and Prognosis

Subacute thyroiditis usually resolves completely and spontaneously over weeks or months. Occasionally, the disease may begin to resolve and then suddenly become worse, sometimes involving first one lobe of the thyroid gland and then the other (migrating or *creeping* thyroiditis). Exacerbations may occur when the FT_4 levels have fallen, the TSH level has risen, and the gland is starting to recover function. Rarely, the course may extend over several years, with repeated bouts of inflammatory disease. Ninety percent of patients recover without any sequelae. In about 10% of patients, permanent hypothyroidism ensues, and long-term T_4 therapy is necessary.

2. CHRONIC THYROIDITIS

Chronic thyroiditis (Hashimoto thyroiditis and lymphocytic thyroiditis) is the most common cause of hypothyroidism and goiter in the United States. It is certainly the major cause of goiter in children and young adults and is the cause of *idiopathic myxedema*, which represents an end stage of Hashimoto thyroiditis, with total destruction of the gland.

Etiology and Pathogenesis

Hashimoto thyroiditis is an immunologic disorder in which lymphocytes become sensitized to thyroidal antigens and autoantibodies are formed that react with these antigens (see thyroid autoimmunity, earlier). In Hashimoto thyroiditis, the three most important thyroid autoantibodies are: TG Ab, TPO Ab,

and *TSH-R*–blocking Ab. During the early phases of Hashimoto thyroiditis, TG Ab is markedly elevated, and TPO Ab is slightly elevated. Later, TG Ab may disappear, but TPO Ab is present for many years. Blocking Abs may be found in patients with atrophic thyroiditis and myxedema and, rarely, in mothers giving birth to infants with no detectable thyroid tissue (**athyreotic cretins**). The pathology of Hashimoto thyroiditis involves a heavy infiltration of lymphocytes totally destroying normal thyroidal architecture. Lymphoid follicles and germinal centers may be formed. The follicular epithelial cells are frequently enlarged and contain an eosinophilic cytoplasm laden with mitochondria (Hürthle cells). Destruction of the gland results in a fall in serum T_3 and FT_4 and a rise in TSH. Initially, TSH may maintain adequate hormonal synthesis by the development of thyroid enlargement or goiter, but often the gland fails, and hypothyroidism with or without goiter ensues.

Hashimoto thyroiditis is part of a spectrum of thyroid diseases that includes Graves disease at one end and idiopathic myxedema at the other (Figure 7–44). It is familial and may be associated with other autoimmune diseases, including pernicious anemia, adrenocortical insufficiency, idiopathic hypoparathyroidism, myasthenia gravis, and vitiligo. *Schmidt syndrome* consists of Hashimoto thyroiditis, idiopathic adrenal insufficiency, and, commonly, type 1 diabetes mellitus. This phenomenon has also been called the autoimmune polyglandular syndrome (see Chapter 2).

Clinical Features

A. Symptoms and signs Hashimoto thyroiditis usually presents with goiter in a patient who is euthyroid or has mild hypothyroidism. The female:male sex distribution is about 4:1. The process is painless, and the patient may be unaware of the goiter unless it becomes very large. Older patients may present with severe hypothyroidism with only a small, firm atrophic thyroid gland (idiopathic myxedema).

B. Laboratory findings There are multiple defects in iodine metabolism. RAIU may be high, normal, or low. Circulating thyroid hormone levels are usually normal or low, and if low, TSH will be elevated.

The most striking laboratory finding is the high titer of autoantibodies to thyroidal antigens in the serum. Serum tests for either TG or TPO Abs are positive in most patients with Hashimoto thyroiditis. Another diagnostic test that may be helpful is FNAB, which reveals lymphocytic infiltration as well as the presence of Hürthle cells.

Idiopathic myxedema	Hashimoto thyroiditis	Euthyroid Graves disease	Graves disease
Hypothyroid without goiter	Hypothyroid or euthyroid with goiter	Euthyroid with or without goiter	Hyperthyroid with goiter

FIGURE 7–44 Spectrum of autoimmune disease of the thyroid gland. The clinical manifestations of autoimmune disease of the thyroid gland range from idiopathic myxedema, through nontoxic goiter, to diffuse toxic goiter, or Graves disease. Progression of autoimmune disease from one form to another in the same patient can occasionally occur.

Differential Diagnosis

Hashimoto thyroiditis can be differentiated from other causes of nontoxic goiter by serum Ab studies and, if necessary, by FNAB.

Complications and Sequelae

The major complication of Hashimoto thyroiditis is progressive hypothyroidism. Most patients with Hashimoto thyroiditis initially have a small goiter and subclinical hypothyroidism, which is defined as normal serum levels of FT_4 and T_3, but mildly elevated serum TSH levels, usually less than 10 mU/L. This is in contrast to overt hypothyroidism, in which FT_4 levels are subnormal. Whether subclinical hypothyroidism is a significant health problem that warrants therapy is a matter of debate. Some patients may have mild symptoms of hypothyroidism as well as increased lipid levels and other risk factors for atherosclerotic cardiovascular disease. There may be progression to overt hypothyroidism over time, especially if serum levels of antithyroid Abs are high. On the other hand, most patients are asymptomatic, especially when TSH serum levels are less than 10 mU/L, and the link between atherosclerosis is still controversial.

One variant of Hashimoto thyroiditis has been termed **silent or painless thyroiditis**. It has a predilection for occurring in the postpartum period, developing in about 5% of postpartum women. In this setting it is called **postpartum thyroiditis**. Most patients have symptoms and signs of mild hyperthyroidism that may be difficult to distinguish from mild Graves disease. There are no eye findings, and unlike subacute thyroiditis, the thyroid is not tender to palpation. The serum FT_4 is elevated out of proportion to the serum T_3, typical of most forms of thyroiditis, when there is leakage of stored hormone into the bloodstream. The erythrocyte sedimentation rate is normal, which distinguishes it from subacute thyroiditis, and TPO Ab titers are elevated. Importantly, the 24-hour RAIU is low, rather than elevated, which readily distinguishes this condition from postpartum Graves disease. Silent thyroiditis may evolve with a triphasic course, similar to what is seen in subacute thyroiditis, with the hyperthyroid phase lasting 1 to 3 months, followed by a hypothyroid phase lasting several months as well. In postpartum thyroiditis, the hyperthyroid phase generally begins 3 to 4 months after delivery. Although recovery is the rule, about 25% of patients have permanent mild or subclinical hypothyroidism, which can progress with time to overt hypothyroidism. Long-term follow-up is required because permanent hypothyroidism can eventually develop in patients after many years. Also, recurrent episodes of silent thyroiditis can occur, and postpartum thyroiditis often develops following subsequent pregnancies.

Rarely, a patient with Hashimoto thyroiditis may develop lymphoma of the thyroid gland. Although the etiology of thyroid lymphoma is unknown, Hashimoto thyroiditis is a definite risk factor. It is possible that thyroid lymphoma may be the result of the expansion of an abnormal clone of an immortalized population of intrathyroidal lymphocytes. Thyroid lymphoma is characterized by rapid growth of the gland despite continued T_4 therapy; the diagnosis of lymphoma can be made by FNAB and immunocytochemical analysis, looking for a monoclonal population of lymphocytes, but sometimes requires surgical biopsy (see later).

Whether adenocarcinoma of the thyroid gland occurs more frequently in patients with Hashimoto thyroiditis is controversial, but the two diseases—chronic thyroiditis and carcinoma—can coexist in the same gland. Recent studies have shown that the presence of thyroid cancer in thyroid nodules is more frequent in patients whose serum TSH is higher, even when it is within the normal range. Cancer must be suspected when a solitary nodule or thyroid mass grows or fails to regress while the patient is receiving doses of T_4 that normalize or suppress the serum TSH. FNAB is critical in the differential diagnosis in this situation.

Treatment

The indications for treatment of Hashimoto thyroiditis are goiter or overt hypothyroidism; a patient with a positive thyroid Ab test alone does not require therapy. Surgery is rarely indicated for Hashimoto thyroiditis but occasionally is performed if a goiter does not regress and continues to cause compressive symptoms. The treatment of subclinical hypothyroidism is a matter of debate but is often instituted because of (1) mild symptoms; (2) dyslipidemia which could be ameliorated by T_4 therapy; and (3) positive antithyroid Ab titers, which predicts a higher chance of progression to overt hypothyroidism over time. Sufficient T_4 is given to normalize TSH and allow regression of the goiter.

The treatment of silent or postpartum thyroiditis varies with the patient's thyroid status. During the hyperthyroid phase, β-adrenergic–blocking drugs can be used to treat the symptoms of tremor, palpitations, and nervousness. T_4 can be used in the hypothyroid phase, but it is usually not needed, because symptoms are usually mild.

Course and Prognosis

Without treatment, Hashimoto thyroiditis usually progresses from goiter and subclinical hypothyroidism to overt hypothyroidism over many years. In severe cases, myxedema and myxedema coma may be the end result of untreated disease. With T_4 therapy, the goiter usually resolves, although not necessarily completely, and hypothyroid symptoms are reversed.

Because Hashimoto thyroiditis may be part of a syndrome of multiple autoimmune diseases (see Chapter 2), the patient should be monitored for other autoimmune diseases such as pernicious anemia, adrenal insufficiency, and type 1 diabetes mellitus. Patients with Hashimoto thyroiditis may also develop true Graves disease, occasionally with severe ophthalmopathy or dermopathy (see Figure 7–44). The chronic thyroiditis may blunt the severity of the thyrotoxicosis, so that the patient may present with eye or skin complications of Graves disease without marked thyrotoxicosis, a syndrome often called **euthyroid Graves disease**. The ophthalmopathy and dermopathy are treated as if thyrotoxic Graves disease were present.

3. OTHER FORMS OF THYROIDITIS

The thyroid gland may be subject to acute abscess formation in patients with septicemia or acute infective endocarditis or from local extension of a pharyngeal infection. Immunocompromised

hosts may develop opportunistic thyroidal infections, with involvement of organisms such as aspergillus, mycobacteria, and pneumocystis. Abscesses cause symptoms of pyogenic infection, with local pain and tenderness, swelling, and warmth and redness of the overlying skin. Needle aspiration confirms the diagnosis and identifies the organism. Treatment includes antibiotic therapy and occasionally incision and drainage. A thyroglossal duct cyst may become infected and present as acute suppurative thyroiditis. This too responds to antibiotic therapy and occasionally incision and drainage.

Riedel thyroiditis is associated with extensive fibrosis extending outside the gland and involving overlying muscle and surrounding tissues. Riedel struma (struma is an old term for the thyroid) presents as a stony-hard mass that must be differentiated from thyroid cancer. It is also associated with fibrosis in other parts of the body including the mediastinum and retroperitoneum. Many experts believe that Riedel thyroiditis is a part of systemic illness termed "Immunoglobulin (Ig) G4-related sclerosing disease (IgG4-RSD)" that is characterized by IgG4+ plasma cell infiltration and fibrosis in one or more organs, including the thyroid gland.

EFFECTS OF IONIZING RADIATION ON THE THYROID GLAND

Ionizing radiation can induce both acute and chronic thyroiditis. Thyroiditis may occur acutely in patients treated with large doses of radioiodine and may be associated with release of thyroid hormones and an acute thyrotoxic crisis. Such an occurrence is extremely rare, however, and pretreatment with antithyroid drugs to bring the patient to a euthyroid state prior to ^{131}I therapy should prevent this complication. Radioiodine therapy, which exposes the thyroid to a dosage of around 10,000 rads (cGy), is not associated with the development of thyroid cancer, presumably because the thyroid gland is largely destroyed by high doses of radiation. Thus, although the incidence of postradiation hypothyroidism is high, the incidence of thyroid cancer is extremely low.

External radiation was used many years ago for the treatment of respiratory problems in the newborn, thought to be due to thymic hyperplasia, and for the treatment of benign conditions such as severe acne and chronic tonsillitis or adenoiditis. This treatment has been associated with the later development of nodular goiter, hypothyroidism, or thyroid cancer. Fortunately, such therapy has not been used since the early 1960s. In addition, high doses of therapeutic radiation in children with Hodgkin disease or other cancers have also been associated with the development of thyroid cancer later in life. Almost all radiation-related thyroid cancers have been of the papillary type. In some parts of the world, radiation exposure from fallout from atomic bomb testing or a nuclear reactor accident has been associated with nodular thyroid disease and thyroid cancer. Although the overall frequency of thyroid carcinoma in irradiated patients is low, data from several large series suggest that the incidence of cancer in a patient who presents with a solitary cold nodule of the thyroid

gland and a history of therapeutic radiation of the head, neck, or chest is about 50%.

The most important episode of large-scale radiation-induced thyroid neoplasia was the Chernobyl disaster in April, 1986, at which time huge amounts of radioactive material, especially radioiodine, were released. As early as 4 years later, a striking increase in the incidence of thyroid nodules and thyroid cancer was noted in children in Gomel, an area in the Republic of Belarus that is close to Chernobyl and was heavily contaminated by the accident. A high proportion of the cancers arose in very young children and developed after a short latency period. The sex distribution was equal. Most of the cancers have been papillary carcinomas but they have been more aggressive than typical papillary cancer, with intraglandular, capsular, local, and lymph node invasion. A smaller release of radioiodine also occurred in 2011 in Fukushima, Japan after an earthquake and tsunami destroyed a nuclear power plant there. Thyroid ultrasound screening of children living in the area began in 2012, but thus far no increase in thyroid neoplasia in the exposed population has been reported.

Patients who have been exposed to ionizing radiation should be followed carefully for life. Annual studies should include physical examination of the neck for goiter or nodules and FT_4 or TSH determinations to rule out hypothyroidism. Periodic thyroid ultrasound may detect nodules that are not palpable. If a nodule is found and is greater than 1 cm in diameter, FNAB should be performed. If the nodule is malignant, the patient should have total thyroidectomy. If it is benign, follow-up with periodic thyroid ultrasound is recommended. If the nodule enlarges, repeat biopsy or surgery should be considered.

THYROID NODULES AND THYROID CANCER

Thyroid nodules are extremely common, particularly among women. The prevalence of palpable thyroid nodules in the United States has been estimated to be about 4% of the adult population, with a female-male ratio of 4:1. In young children, the prevalence is less than 1%; in persons aged 11 to 18 years, about 1.5%; and in persons over age 60, about 5%. In contrast to the relatively low prevalence of palpable nodules, thyroid nodules are detected on thyroid ultrasound in up to 50% of healthy individuals, especially middle-aged and older women.

Although thyroid nodules are common, thyroid cancer is a relatively rare condition, with an annual incidence of approximately 8.7 per 100,000 population. The incidence of diagnosed thyroid cancer has increased dramatically over the last two decades, in large part due to the discovery of thyroid nodules and cancers incidentally on imaging done for other reasons, including cervical CT scans and MRI, PET scans, and carotid Doppler testing. In 2017, there are expected to be approximately 57,000 new cases in the United States, compared to 37,000 in 2009. While the vast majority (90–95%) of thyroid nodules are benign, and it is important to identify those that are likely to be malignant in the most rational and cost-effective manner. In 95% of cases,

thyroid cancer presents as a nodule or lump in the thyroid, found on routine physical examination or incidentally on imaging for another reason. Occasionally, especially in children, enlarged cervical lymph nodes are the first sign of the disease, although on careful examination a small thyroid nodule can often be felt. Rarely, distant metastasis in lung or bone is the first sign of thyroid cancer.

1. BENIGN THYROID NODULES

Etiology

Benign conditions that can produce nodularity in the thyroid gland are listed in Table 7–10. They include focal areas of chronic thyroiditis; colloid-filled or hyperplastic cellular nodules within a multinodular goiter; a cyst involving thyroid tissue; rarely parathyroid tissue or thyroglossal duct remnants; and agenesis of one lobe of the thyroid, with hypertrophy of the other lobe presenting as a mass in the neck and mimicking a nodule. It is usually the left lobe of the thyroid that fails to develop, and the hypertrophy occurs in the right lobe. Finally, benign areas of hyperplasia and neoplasms in the thyroid, including follicular adenomas and Hürthle cell adenomas (also called oxyphil adenomas) present as thyroid nodules. Rare benign thyroidal lesions include teratomas, lipomas, and hemangiomas.

Differentiation of Benign and Malignant Lesions

Risk factors that predispose to benign or malignant disease are set forth in Table 7–11 and discussed later.

A. History   A family history of goiter suggests benign disease, as does residence in an area of endemic goiter. However, a family history of papillary or medullary carcinoma or a history of recent thyroid growth, hoarseness, dysphagia, or obstruction suggests cancer, as does exposure to ionizing radiation, as discussed earlier.

B. Clinical features   Clinical features associated with a lower risk of thyroid cancer include middle age, female sex, a soft thyroid nodule on physical examination, and the presence of a multinodular goiter. Individuals at higher risk of thyroid cancer include children, young adults, and older males. A solitary firm

TABLE 7–10   Etiology of benign thyroid nodules.

1. Focal thyroiditis
2. Dominant nodule in a multinodular goiter
3. Benign adenomas:
   a. Follicular
   b. Hürthle cell
4. Thyroid, parathyroid, or thyroglossal cysts
5. Agenesis of a thyroid lobe
6. Postsurgical remnant hyperplasia
7. Postradioiodine remnant hyperplasia
8. Rare: teratoma, lipoma, hemangioma

or dominant nodule that is clearly different from the rest of the gland signifies an increased risk of malignancy. Vocal cord paralysis, enlarged cervical lymph nodes, and suspected metastases are strongly suggestive of malignancy.

C. Serum factors   A serum TSH level should be the initial laboratory study in the evaluation of a thyroid nodule. The serum TSH level is an independent risk factor for thyroid cancer, with the risk of cancer rising progressively with serum TSH levels still within the normal range. For example, in one meta-analysis, the odds ratio for thyroid cancer in a nodule rose by 1.16 for each mU/L of TSH above 1 mU/L. Furthermore, higher serum TSH values have also been associated with more advanced degrees of thyroid cancer. If the serum TSH level is normal, which is the case for most patients, no further blood tests are needed. If the serum TSH is below normal, a radionuclide thyroid scan should be considered to see whether the nodule is hot or autonomously functioning (see later). In the presence of a firm thyroid gland or other clinical features of Hashimoto thyroiditis (eg, positive family history of hypothyroidism), serum anti-TG and anti-TPO Abs should be assessed. A high titer of thyroid autoantibodies in serum suggests chronic thyroiditis as the cause of thyroid enlargement and nodules, but does not rule out an associated malignancy. Measurement of serum calcitonin is not routinely advised. However, an elevated serum calcitonin in patients with a family history of medullary carcinoma strongly suggests the presence of medullary thyroid cancer. Elevated serum TG following total thyroidectomy for papillary or follicular thyroid cancer usually indicates residual or metastatic disease, but serum TG is not helpful in determining the nature of a thyroid nodule.

D. Imaging studies   Thyroid ultrasound has become the most important tool in the confirmation and differential diagnosis of thyroid nodules and provides considerably more information than radionuclide thyroid scans (see Figure 7–27). Ultrasound is used to confirm that a palpable neck mass is actually a thyroid nodule; to measure it establishes a nodule's size precisely and, hence, can be used for subsequent monitoring; it identifies specific sonographic features that are important in assessing the possibility of malignancy; and it often detects other nonpalpable nodules elsewhere in the gland that are clinically significant. In fact, the frequency of malignancy is the same whether a nodule is palpable or only detectable with imaging.

The sonographic features that are associated with a higher risk of malignancy include microcalcifications, irregular borders, and a "taller than wide" configuration (Table 7–12 and Figure 7–45). Other ultrasound characteristics, such as hyperechogenicity, a thin sonolucent rim ("halo") around the nodule, or a large cystic component favor a benign nodule (see Figure 7–45).

In patients with low serum TSH levels, radionuclide thyroid scanning with [123]Iodine or [99m]Technetium pertechnetate can be used to identify whether the nodule is hot or cold (ie, a nodule that takes up more [hot] or less [cold] radioactive material than surrounding thyroid tissue). Hot nodules are almost never

TABLE 7–11   Risk factors useful in distinguishing benign from malignant thyroid lesions.

	More Likely Benign	More Likely Malignant
History	Family history of benign goiter Residence in endemic goiter area	Family history of medullary cancer of thyroid Previous therapeutic irradiation of head or neck Recent growth of nodule Hoarseness, dysphagia, or obstruction
Physical characteristics	Older woman Soft nodule Multinodular goiter	Child, young adult, male Solitary, firm nodule clearly different from rest of gland (dominant nodule) Vocal cord paralysis, firm lymph nodes, distant metastases
Serum factors	High titer of thyroid autoantibodies	Elevated serum calcitonin
Scanning techniques		
^{123}I or ^{99m}TcO$_4$	Hot nodule	Cold nodule
Sonogram	Cyst (pure)	Solid or complex
Biopsy (needle)	Benign appearance on cytologic examination	Malignant or suggestion of malignancy
Levothyroxine therapy (TSH suppression for 3-6 mo)	Regression	No regression

malignant, and although 5% to 10% of cold nodules may be malignant, 90% to 95% of cold nodules are benign (Figure 7–46). On the other hand, thyroid scans are rarely helpful in patients with normal or elevated serum TSH levels.

E. Needle biopsy   FNAB of thyroid nodules to obtain cytological material for pathological examination is the principal test to distinguish benign from malignant lesions. Targeting thyroid nodules with suspicious ultrasound features for biopsy provides a higher yield of thyroid cancer than aspirating all nodules larger than 1 cm in diameter. Recent guidelines provide a framework for decision-making, which incorporates sonographic features plus nodule size (Table 7–13; see Figure 7–45). Purely cystic nodules without a mural component do not require biopsy, but the fluid could be aspirated if the nodule is causing pressure symptoms because of its size. However, cystic nodules often recur following aspiration.

The technique of FNAB is relatively simple, safe, reliable, and well tolerated. FNAB is now generally performed under sonographic guidance and definitely should be done so if the nodule

TABLE 7–12   Ultrasound characteristics of thyroid nodules.

Characteristics of Benign Thyroid Nodules
Iso-Hyperechoic
Spongiform appearance
Smooth border or "halo"
Thin wall cyst
Dense calcifications

Characteristics of Malignant Thyroid Nodules
Microcalcifications
Irregular borders
Hypoechoic
Increased intranodular vascularity
Taller than wide

is nonpalpable, difficult to localize by palpation, or is cystic. FNAB separates thyroid nodules into six groups (Table 7–14 and Figure 7–47):

1. Nondiagnostic (insufficient).
2. Cytologically benign thyroid nodules (70%-80% of all biopsy specimens), which exclude cancer with a negative predictive value of 95% to 98%. These include adenomatoid nodules, colloid nodules, and Hashimoto thyroiditis.
3. Follicular lesion or atypia of undetermined significance (FLUS or AUS). This category includes nodules with nuclear atypia and those with mixed macro- and microfollicular architecture, those predominantly comprised of Hürthle cells.
4. Follicular neoplasm: This category includes microfollicular nodules and Hürthle cell nodules (Hürthle cells are large eosinophilic cells derived from follicular epithelium also called oxyphils).
5. Suspicious for malignancy: Specimen contains some cytologic features typically associated with cancer, but not enough to be confidently diagnosed as cancer.
6. Malignant: Cytologically malignant thyroid nodules (3%-5% of all biopsy specimens), which have a positive predictive value for cancer of 95% for all types of thyroid malignancies.

Nodules that are FLUS/AUS, or follicular neoplasms are grouped together as "indeterminate" or "suspicious." However, the majority of such nodules are benign.

Until recently, patients with indeterminate thyroid nodules on FNAB were sent to surgery for definitive diagnosis, which, in retrospect was often unnecessary, as the majority of indeterminate nodules are benign on histopathology. Recently, however, new tools have become available to assist in refining the diagnosis and to decrease the number of unnecessary surgeries. These "molecular tests" take advantage of recently acquired knowledge about the molecular genetics of thyroid neoplasia (Figure 7–48).

Gene point mutations and rearrangements are found in over 75% of thyroid cancers. For papillary thyroid cancer, mutations in genes coding for proteins in the MAPK signaling pathways,

FIGURE 7–45 American Thyroid Association categorization of sonographic features of thyroid nodules from benign/low suspicion to very high suspicion. (Reproduced with permission from Haugen BR Md, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016 Jan;26(1):1-133.)

which regulate cell proliferation, differentiation, and survival, are of greatest importance. These include point mutations in BRAF (the V600E mutation), occurring in about 40% of papillary cancers, and *RAS* gene mutations (N-RAS, H-RAS, and K-RAS), which occur in about 20% of papillary cancers, in 30% to 40% of follicular cancers, and in approximately 20% of benign follicular adenomas. Additional genetic abnormalities seen in papillary thyroid cancers include gene rearrangements in which a portion of the *RET* gene, encoding a receptor tyrosine kinase not normally expressed in the thyroid, is linked to a portion of one of several unrelated genes at its 5′ end, producing a constitutively active chimeric RET receptor, collectively referred to as RET/PTC oncogenes. These oncogenic genetic changes are linked to radiation exposure. Finally, about 30% of follicular carcinomas, and a smaller fraction of follicular adenomas, contain a *PAX8/PPARγ* gene fusion, in which PAX8, a thyroid transcription factor is fused to the peroxisome proliferator-activated receptor, gamma isoform.

Based on these observations, biopsy specimens with indeterminate cytology can be subjected to further "molecular" analysis in an effort to increase the diagnostic accuracy and avoid unnecessary surgery. Two strategies are currently available commercially. In one method, using a so-called "mutation panel," DNA from the cells in the aspirate is analyzed for a specific mutations in BRAF and RAS, as well as for the presence of RET/PTC rearrangments and the PAX8/PPAR fusion. The presence of one of these genetic abnormalities is associated with a high-positive predictive value for malignancy (85%-90%). However, currently available mutation panels do not routinely detect mutations in 10% to 15% of patients with proven malignancy, and therefore a negative test does not confidently rule out the possibility of thyroid cancer. Newer mutational testing strategies, which include other genes in the panel, including TERT (telomerase reverse transcriptase) and the *TP53* tumor suppressor gene, may have even better sensitivity, but have not been validated.

The *mRNA* gene expression classifier is the other testing strategy that can be performed on indeterminate thyroid aspirates. In this method, the expression of a panel of 167 genes within the nodule is assessed by measuring mRNA levels using microarray technology. Validation studies of this method suggest that a
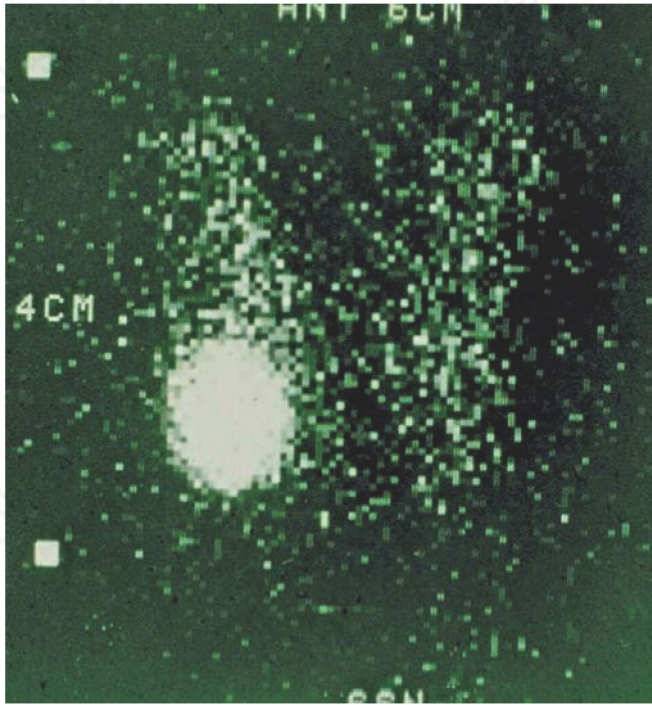
FIGURE 7-46 A functioning thyroid nodule in a patient with a palpable right lower pole thyroid nodule and a low serum TSH concentration but normal serum-Free T_4 and T_3 levels.

"benign" result is associated with a 94% to 95% negative predictive value, which would suggest that surgery can be postponed and the nodule observed. While most patients with suspicious gene classifier results opt for surgery, up to half of them will have a benign nodule on final pathology. Another strategy that is under development is the assessment of the expression of specific microRNAs (miRNAs) that are closely linked to thyroid malignancy.

Management of Thyroid Nodules

A decision tree for management of a thyroid nodule is presented in Figures 7–49A and 7–49B. A patient with a thyroid nodule, detected either on physical examination or imaging, should have a serum TSH and thyroid ultrasound performed. If the TSH is low, radionuclide thyroid scanning should be done to avoid performing FNAB in a patient with a hot nodule. If the serum TSH is normal or elevated, and ultrasound confirms the presence of one or more nodules, FNAB should be performed on those hypoechoic nodules that are larger than 1.0 to 1.5 cm, especially in the presence of other suspicious sonographic characteristics. If the nodule is cytologically malignant or if the reading is "suspicious for malignancy," the patient is generally referred for thyroid surgery unless there are other serious medical problems associated with high-operative risk and/or a prognosis much graver than thyroid cancer. If the nodule is cytologically benign and neither local compressive symptoms nor hyperthyroidism are present, only reassessment of nodule size in 6 to 18 months by physical examination and/or ultrasound is required. If the nodule grows, defined as an increase of 20% or more in two of three dimensions, repeat FNAB and cytological examination may be indicated to verify the benign diagnosis. However, in this situation, the original benign results are confirmed 98% to 99% of the time.

In patients with cytologically indeterminate nodules, a radionuclide scan may be obtained if the serum TSH is at the lower end of the normal range. If the scan reveals the nodule to be hot (ie, functioning and suppressing the remainder of the gland), the patient can be observed to detect future overt hyperthyroidism. If the lesion is cold and in all other situations where the serum TSH is within the reference range, both surgery (a diagnostic

TABLE 7–13 Sonographic patterns, estimated risk of malignancy, and fine-needle aspiration guidance for thyroid nodules.

Sonographic Pattern	Sonographic Features	Estimated Risk of Malignancy (%)	Size Cutoff for Biopsy
High suspicion	Solid hypoechoic nodule or solid hypoechoic component of a partially cystic nodule with one or more of the following features: irregular margins (infiltrative, microlobulated), microcalcifications, taller than wide shape, rim calcifications with small extrusive soft tissue component, evidence of ETE	>70-90	Recommend FNA at ≥1 cm
Intermediate suspicion	Hypoechoic solid nodule with smooth margins without microcalcifications, ETE, or taller than wide shape	10-20	Recommend FNA at ≥1 cm
Low suspicion 5-10	Isoechoic or hyperechoic solid nodule, or partially cystic nodule with eccentric solid areas, without microcalcification, irregular margin or ETE, or taller than wide shape	5-10	Recommend FNA at ≥1.5 cm
Very low suspicion	Spongiform or partially cystic nodules without any of the sonographic features described in low, intermediate, or high suspicion patterns	<3	Consider FNA at ≥2 cm Observation without FNA is also a reasonable option
Benign	Purely cystic nodules (no solid component)	<1	No biopsy

Abbreviation: ETE, Extrathyroidal Extension.

Reproduced with permission from Haugen BR Md, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid.* 2016 Jan;26(1):1-133.

TABLE 7–14 **The Bethesda system for reporting thyroid cytopathology[*].**

Bethesda Class	Diagnostic Category	Predicted Cancer Risk (%)	Actual Cancer Risk (% median, range) Based on Meta-analysis by Bongiovanni et al. 2012[**])
1	Nondiagnostic (unsatisfactory) (5%)	1-4	20% (9-32%)
2	Benign (66%)	0-3	2.5% (0-10%)
3	Atypia of undetermined significance (AUS) or Follicular Lesion of Undetermined Significance (FLUS)	5-15	14% (6-48%)
4	Follicular lesion (or suspicious for follicular lesion)	15-30	25% (15-34%)
5	Suspicious for malignancy	60-75	70% (53-97%)
6	Malignant (5%)	97-99	99% (94-100%)

*Adapted with permission from Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for reporting thyroid cytopathology: a meta-analysis. *Am J Clin Pathol*. 2009 Nov;132(5):658-665.

**Data from Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol*. 2012;56(4):333-339.
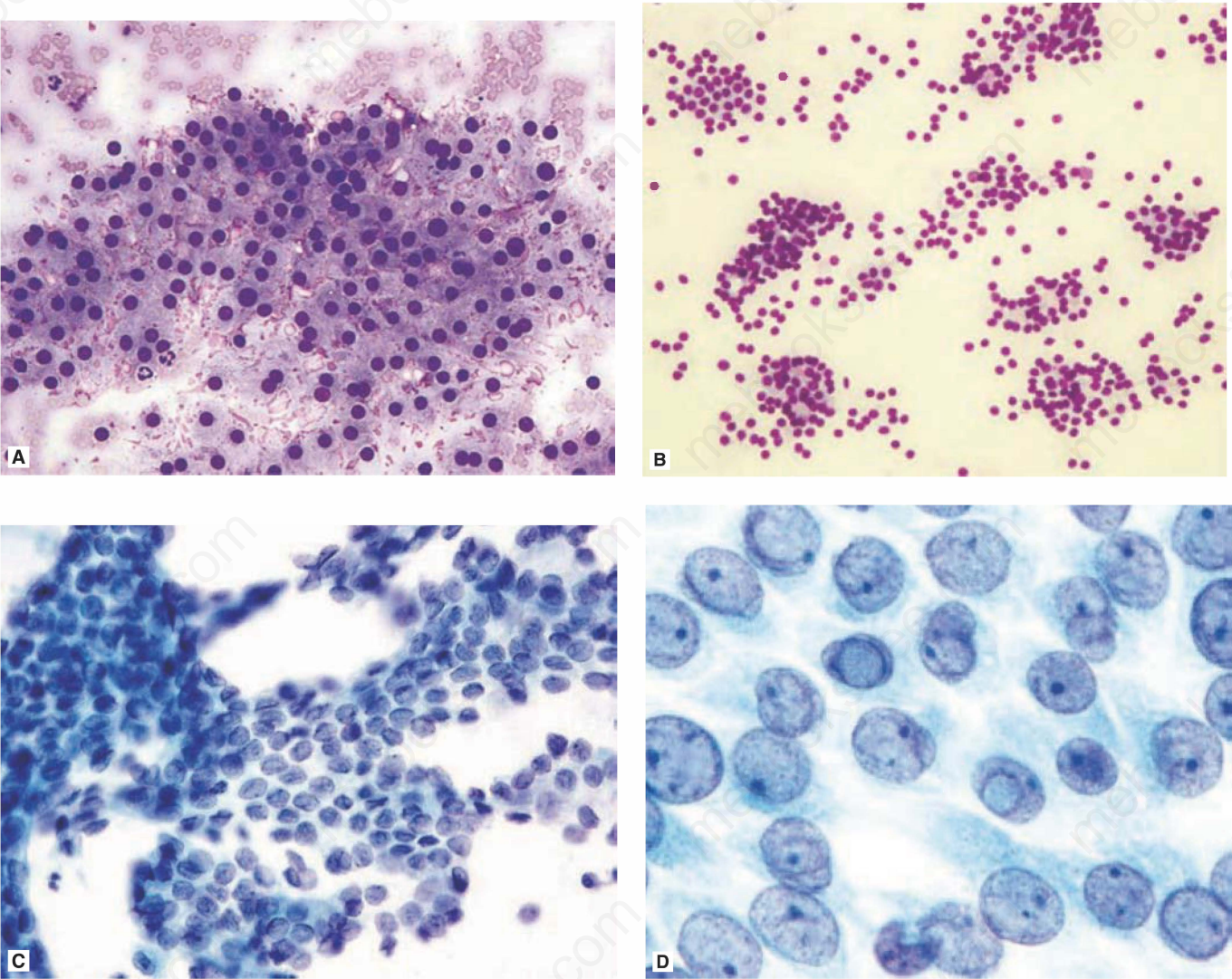
FIGURE 7-47 Representative Thyroid Cytology. **A.** Benign thyroid cells in a monolayer sheet with round nuclei and abundant cytoplasm. **B.** Follicular Neoplasm: Cells have a microfollicular arrangement with overlapping nuclei. **C.** Papillary Thyroid Cancer: Monolayer of cells in a syncytial arrangement with large oval nuclei. **D.** Papillary Thyroid Cancer (higher power) illustrating nuclear inclusions ("orphan Annie eye nuclei"). (Reproduced with permission from Ali SZ, Cibas ES. *The Bethesda System for Reporting Thyroid Cytopathology*. New York: Springer Verlag; 2010.)
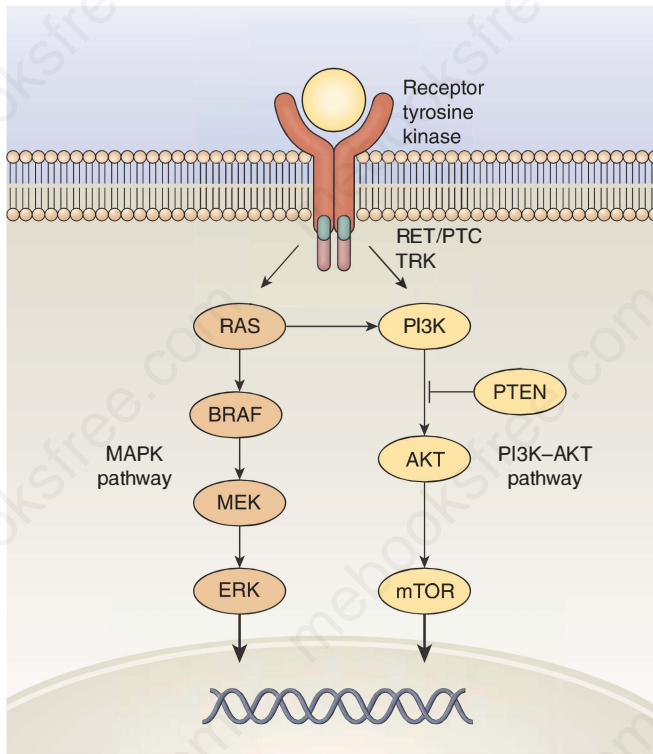
FIGURE 7-48 Signaling pathways for thyroid cell proliferation, differentiation, and survival. Several pathways are involved in propagation of signals from various cell membrane receptor tyrosine kinases into the nucleus. They regulate multiple cell processes including proliferation, differentiation and survival. Activation of the MAPK pathway by oncogenic stimuli such as mutated BRAF, RAS or the chimeric fusion proteins RET/PTC, and TRK is a common tumor initiating event in well-differentiated papillary carcinoma and in some follicular carcinomas. Mutations involving the effectors of the PI3K–AKT pathway such as the PI3K subunit PIK3CA, AKT1, and PTEN are found more frequently in follicular carcinomas and in less differentiated types of thyroid cancer. (Reproduced with permission from Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. *Nat Rev Endocrinol*. 2011 Aug 30;7(10):569-580.)

lobectomy) and molecular testing are reasonable options, depending on patient preference, other clinical features such as patient age, and the nodule's size and sonographic features. On the other hand, if an indeterminate nodule is less than 1 cm in diameter, observation with serial sonography is reasonable, with surgery only if the lesion grows.

2. THYROID CANCER

Pathology

The types and approximate frequency of malignant thyroid tumors are listed in Table 7–15.

A. Papillary carcinoma Papillary carcinoma of the thyroid gland typically presents as a nodule that is firm and solid on thyroid ultrasound, sometimes with internal calcifications. In multinodular goiter, the cancer is usually a dominant nodule—larger, firmer, and/or with ultrasound characteristics more suspicious than other nodules in the gland. About 10% of papillary carcinomas, especially in children, present with enlarged cervical nodes. Careful examination then often also reveals a nodule in the thyroid. Sometimes, there is hemorrhage, necrosis, and cyst formation in the malignant nodule, but on thyroid ultrasound of these lesions, a defined solid component is an indication for FNAB of this nodule component. Finally, papillary carcinoma may be found incidentally as a microscopic focus of cancer in a gland removed for other reasons, such as another indeterminate nodule, multinodular goiter, or Graves disease. Microscopically, papillary cancer typically consists of single layers of thyroid cells arranged in vascular stalks, with papillary projections extending into microscopic cyst-like spaces. The nuclei of the cells are large and pale and frequently contain clear, glassy intranuclear inclusion bodies. About 40% of papillary carcinomas form laminated calcified spheres—often at the tip of a papillary projection—called psammoma bodies. Most papillary thyroid cancers grow slowly and remain confined to the thyroid gland and local lymph nodes. However, these cancers can exhibit intraglandular metastasis and lymph node spread. In some children with papillary thyroid cancer, there are extensive lymph node and pulmonary micrometastases. In older patients, papillary cancer can be aggressive and invade locally into adjacent muscles, nerves, and trachea. In later stages, these tumors may spread to the lung, bone, brain, and other organs. Death is usually due to local disease, with invasion of deep tissues in the neck; less commonly, death may be due to extensive distant metastases. In some older patients, a longstanding, slowly growing papillary carcinoma begins to grow rapidly and converts to undifferentiated or anaplastic carcinoma.

Most papillary carcinomas secrete TG, which can be used as a marker for recurrence or metastasis of the cancer. Some papillary cancers concentrate radioiodine, albeit less efficiently than normal thyroid tissue. In such patients, radioiodine is an effective tool for identification and treatment of extrathyroidal disease.

B. Follicular carcinoma Follicular carcinoma is characterized histologically by the presence of small follicles with poor colloid formation. Follicular carcinoma is distinguished from benign follicular adenoma microscopically by the presence of capsular and/or vascular invasion within the tumor capsule. When only the capsule is involved, follicular cancer is categorized as minimally invasive, a tumor that typically has an indolent course similar to low-grade papillary cancers. When capsular blood vessels are also invaded, follicular cancers are categorized as invasive and have the potential for distant metastatic spread to bone, lung, and other sites. These tumors often retain the ability to concentrate radioactive iodine, form TG, and, rarely, synthesize T_3 and T_4. This may render these tumors responsive to radioactive iodine therapy. When death occurs, it is attributable to local extension or distant metastases with extensive involvement of bone, lungs, and viscera.

A variant of follicular carcinoma is the **Hürthle cell or oxyphil cell carcinoma**, characterized by large individual cells with
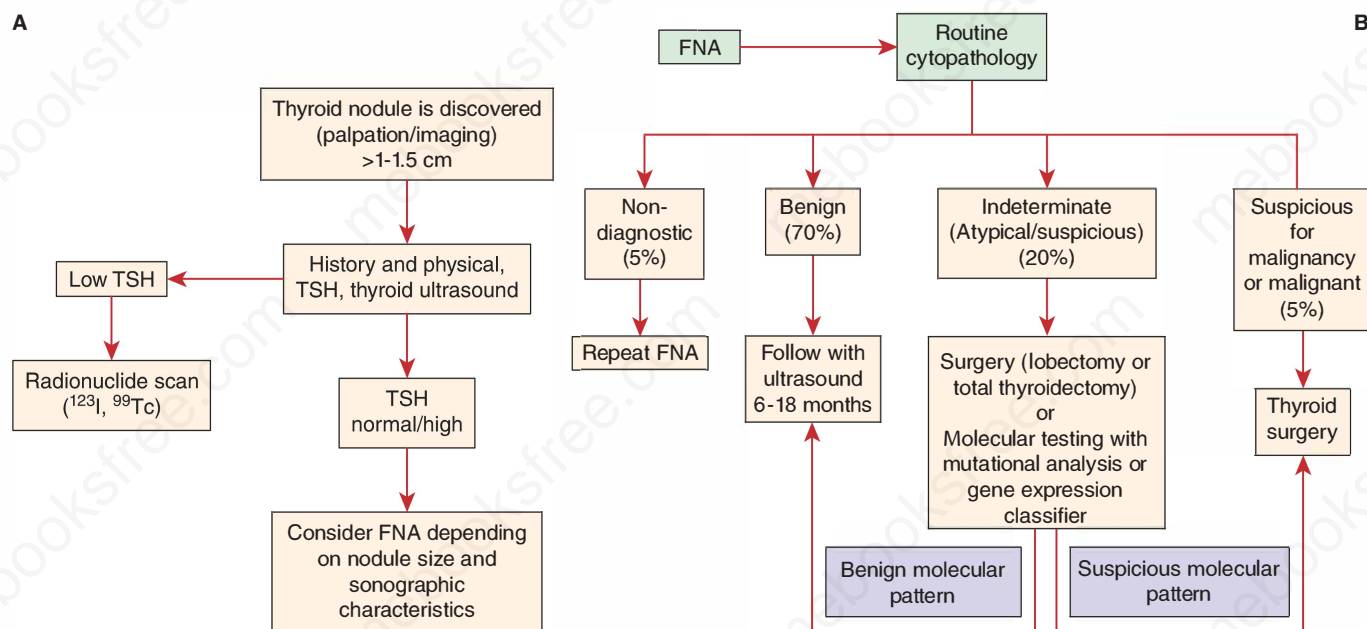
FIGURE 7-49 **A.** Decision matrix for the initial evaluation of a thyroid nodule. All patients with suspected thyroid nodules should have serum TSH measured and a thyroid ultrasound to document the nodule and its sonographic characteristics. If the serum TSH is low, a radionuclide thyroid scan should be considered. **B.** Decision matrix following thyroid fine-needle aspiration (FNA). Nodules that are indeterminate can be managed with surgery (lobectomy or thyroidectomy) or additional evaluation with molecular testing can be undertaken.

pink-staining cytoplasm filled with mitochondria. They behave like follicular cancer except that they rarely take up radioiodine.

C. Medullary carcinoma   Medullary thyroid cancer is a disease of the C cells (parafollicular cells) derived from the ultimobranchial body. These cells are capable of secreting calcitonin, carcinoembryonic antigen (CEA), histaminase, prostaglandins, serotonin, and other peptides (see Chapters 8 and 22). Microscopically, these tumors consist of sheets of cells separated by a pink-staining substance that stains with Congo red, typical of amyloid.

Medullary thyroid carcinoma is typically more aggressive than papillary or follicular carcinoma, extending locally to cervical lymph nodes and into surrounding tissues. It may also invade lymphatics and blood vessels and metastasize to lungs, liver, and other viscera. The calcitonin and CEA secreted by the tumor are clinically useful markers for diagnosis and follow-up.

TABLE 7–15 Approximate frequency of malignant thyroid tumors.

Papillary carcinoma (including mixed papillary and follicular)	80%
Follicular carcinoma (including Hûrthle cell carcinoma)	10%
Medullary carcinoma	5%
Undifferentiated (anaplastic) carcinomas	3%
Miscellaneous (including lymphoma, fibrosarcoma, squamous cell carcinoma, malignant hemangioendothelioma, teratomas, and metastatic carcinomas)	1%

Eighty percent of medullary carcinomas are sporadic, and the remainder are familial. There are three familial patterns: (1) familial medullary cancer without associated endocrine disease; (2) multiple endocrine neoplasia 2A (MEN 2A), consisting of medullary thyroid carcinoma, pheochromocytoma, and hyperparathyroidism; and (3) MEN 2B, consisting of medullary thyroid cancer, pheochromocytoma, multiple mucosal neuromas, as well as intestinal ganglioneuromatosis. There is also a variant of MEN 2A with cutaneous lichen amyloidosis, a pruritic skin lesion located on the upper back. These familial syndromes are due to *RET* protooncogene mutations, most commonly in exons 10, 11, or 16 (see later, Chapters 8 and 22).

If medullary thyroid cancer is diagnosed by FNAB or at surgery, it is essential that the patient be screened for one of the familial medullary thyroid carcinoma syndromes by DNA analysis for mutations in the *RET* protooncogene. If screening is negative, the tumor is almost certainly sporadic, and no investigation of family members is required. If a mutation is found, family members should then be screened.

D. Undifferentiated (anaplastic) carcinoma   Undifferentiated thyroid gland cancer include small cell, giant cell, and spindle cell carcinomas. They usually occur in older patients, who often have a long history of goiter with sudden gland enlargement over weeks or months, causing pain, dysphagia, hoarseness, and/or dyspnea. Death from aggressive local extension and metastatic disease typically occurs within 6 to 36 months. These tumors are very resistant to all currently available therapies.

E. Miscellaneous types

1. **Lymphoma**—Thyroid lymphoma may develop as part of a generalized lymphoma or may be primary in the thyroid gland. Primary thyroid lymphoma usually develops in patients with longstanding Hashimoto thyroiditis. It may be difficult to distinguish from chronic thyroiditis histologically. There is a monoclonal lymphocyte population with invasion of thyroid follicles and blood vessel walls. If there is no systemic involvement, the tumor may respond dramatically to combined radiation and chemotherapy.

2. **Cancer metastatic to the thyroid**—Systemic cancers can metastasize to the thyroid gland. Common primary cancers responsible for these metastases include breast, kidney, lung carcinomas, and malignant melanoma. The primary site of involvement is usually obvious. Occasionally, the diagnosis is made by biopsy of an enlarging cold thyroid nodule. The prognosis is that of the primary tumor.

F. Molecular biology of thyroid neoplasms   Extensive studies have revealed frequent gene mutations in both benign and malignant thyroid neoplasms (Figures 7–48 and 7–50). Activating mutations in the *gsp* oncogene, encoding Gsα or the TSH receptor in thyroid cells have been associated with increased growth and function in benign hot or toxic thyroid nodules. Activating mutations of the *RAS* oncogene and loss of function of *PTEN* a tumor suppressor gene, are common in benign follicular adenomas. Progression to follicular carcinoma may occur after another genetic event involving a chromosomal translocation forming a fusion gene between the thyroid transcription factor *PAX8* gene and the *PPARγ* gene (*PAX8-PPARγ*). Further loss of the suppressor gene *P53* may allow progression to an anaplastic carcinoma. Papillary cancer is likely caused by mutations in genes in the MAP kinase–signaling pathway, especially *BRAF*. Subsequent loss of *P53* suppressor gene may allow progression to anaplastic carcinoma. As noted earlier, activating

mutations of the *RET* protooncogene on chromosome 10 have been shown to be associated with MEN 2A, MEN 2B, and familial medullary thyroid cancer. About 85% to 90% of the mutations found in MEN 2A and familial medullary thyroid cancer occur in exons 10 and 11, whereas about 95% of the mutations associated with MEN 2B are found in codon 918 in exon 16 of the *RET* oncogene. These mutations can be demonstrated in DNA from peripheral white blood cells utilizing the polymerase chain reaction and DNA sequencing. In patients with MEN 2 or familial medullary thyroid cancer who do not demonstrate a mutation in *RET*, family unit linkage analysis may be used to identify gene carriers. Thus, families can be screened for the carrier state, and early diagnosis and treatment can be instituted (Figure 7–51). Somatic mutations in the *RET* oncogene occur in about 30% of sporadic medullary thyroid cancer cells, but this does not occur in white blood cells and does not represent a germline mutation (see also Chapter 22).

Management of Thyroid Cancer (Figure 7–52)

A. Papillary and follicular carcinoma: staging and treatment   Staging of cancer usually relies on the tumor-nodes-metastases (TNM) system for staging thyroid cancer (Table 7–16). Papillary and follicular thyroid carcinomas are grouped together, and staging is related to the age of the patient at the time of diagnosis. The cause-specific 5-year mortality rates are for stage 1, 0%; stage 2, 1%; stage 3, 5%; and stage 4, 77%. The TNM system may underestimate the risk of recurrence and death in younger patients with aggressive disease. Also, the TNM system was designed to predict cancer-related mortality, not the risk of recurrence. Since most patients with differentiated thyroid cancer survive their illness, but their course may be punctuated with recurrent disease over time that may require
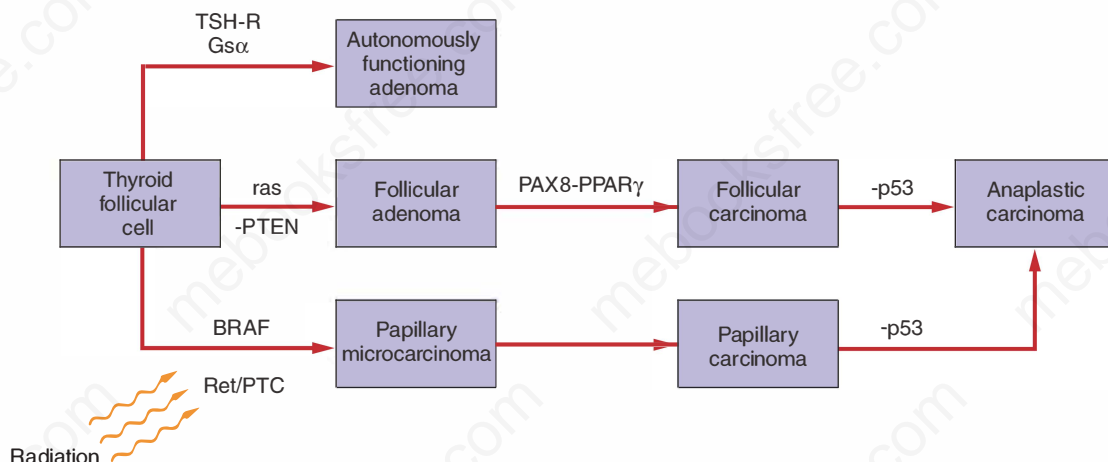
FIGURE 7–50 Molecular defects associated with development and progression of human thyroid neoplasms. The hypothetical role of specific mutational events in thyroid tumorigenesis is inferred from their prevalence in the various thyroid tumor phenotypes (see text for details). (Reproduced with permission from Fagin JA. Genetic basis of endocrine disease 3: molecular defects in thyroid gland neoplasia. *J Clin Endocrinol Metab.* 1992 Dec;75(6):1398-1400.)
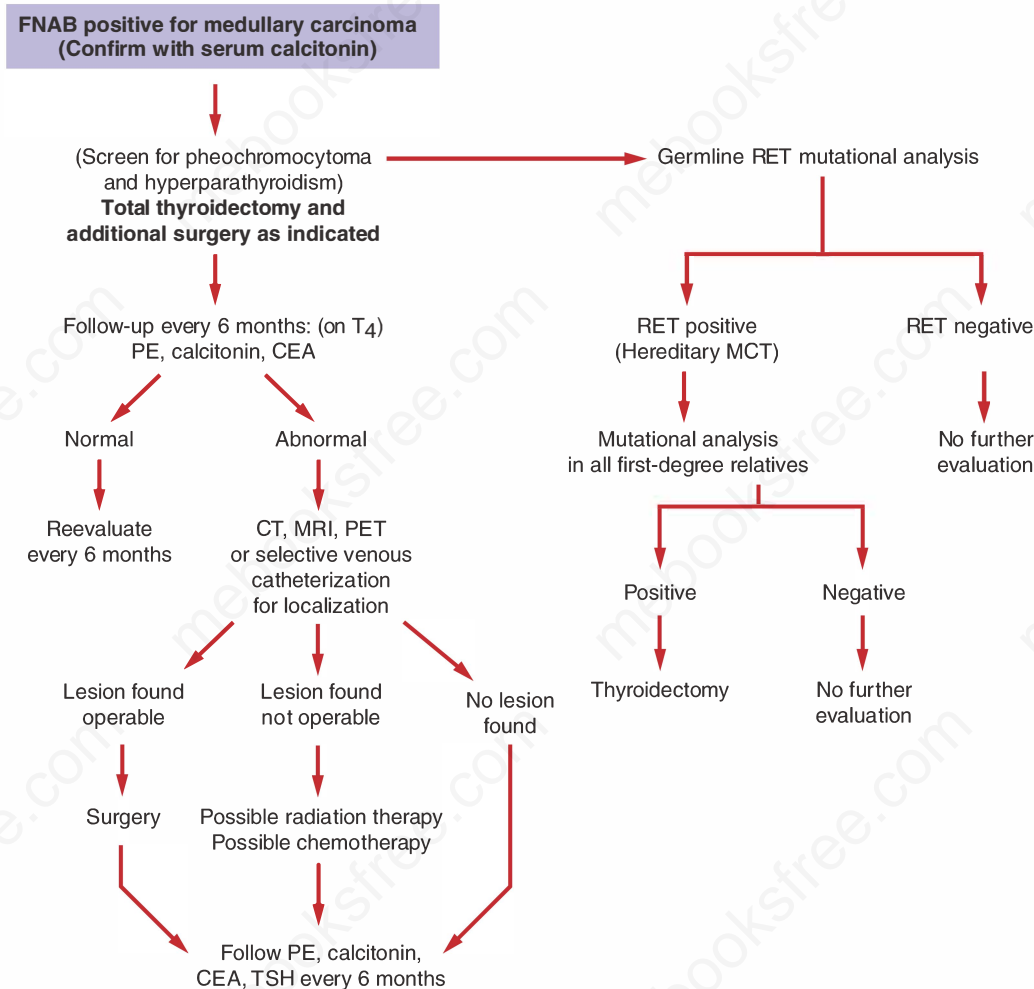
FIGURE 7–51 Decision matrix for the management of medullary thyroid carcinoma (see text for details). (FNAB, fine-needle aspiration biopsy; PE, physical examination.)

therapy with additional surgery or radioiodine, the American Thyroid Association has developed a validated risk stratified system that is able to predict recurrent disease (Table 7–17).

Because both papillary and follicular cancers arise from thyroid epithelial cells and sometimes respond to TSH and concentrate iodine, they are grouped together as **differentiated thyroid cancers**. These patients may be classified loosely into low-risk and high-risk groups. The low-risk group includes patients under age 45 with primary lesions under 4 cm and no evidence of intra- or extraglandular spread. For these patients, lobectomy is probably adequate therapy. All other patients should be considered to be at higher risk, and for these, total thyroidectomy and—if there is preoperative or intraoperative evidence of lymphatic spread—central compartment neck dissection are indicated. In the absence of evidence of lymphatic spread, prophylactic neck dissection may not be necessary, but can be considered for larger tumors (>4 cm).

For some low-risk patients, especially those under age 45 with large or numerous lymph node metastases, and all high-risk patients, postoperative radioiodine ablation of residual thyroid

remnant is recommended as adjuvant therapy in an effort to decrease the likelihood of recurrent disease. To achieve maximal uptake of radioiodine into the remnant, an elevated serum TSH level is required. This can be achieved using injections of recombinant human TSH (rhTSH) given to patients in the euthyroid state. Alternatively, TSH stimulation can be achieved by withdrawing the patient from T_4 therapy, which causes transient clinical hypothyroidism. If the latter method is employed, T_3 is often prescribed temporarily at a dose of 25 to 50 µg daily in divided doses for 2 to 4 weeks; the medication is then stopped for 2 weeks, and, because of its 1-day half-life, patients quickly become hypothyroid prior to radioiodine therapy. In general, rhTSH is preferred, since clinical hypothyroidism is avoided and whole body radiation exposure is lower as the radioiodine is cleared from the body more rapidly because renal function is preserved. Regardless of how the patient achieves an elevated serum TSH level, a low-iodine diet is prescribed for 1 to 2 weeks to enhance uptake of the radioisotope. Following either rhTSH stimulation or thyroid hormone withdrawal, the serum TG level is
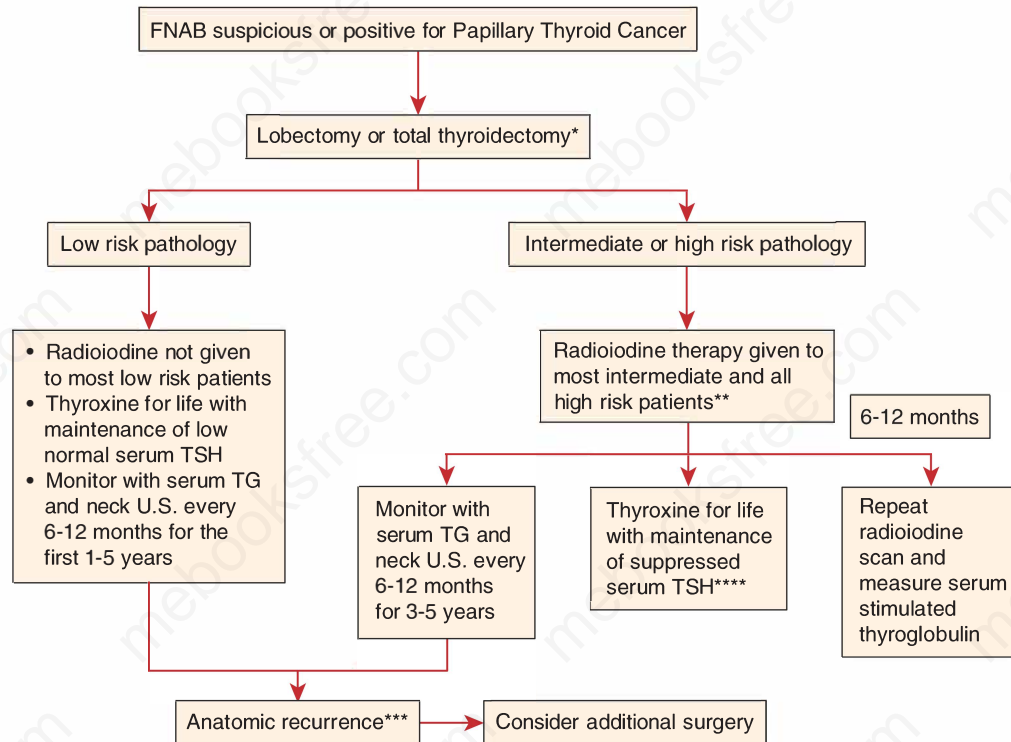
FIGURE 7–52 Decision matrix for the management of papillary or follicular thyroid cancer based on 2015 American Thyroid Association Guidelines. (Data from Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016 Jan;26(1):1-133.)

*Either lobectomy or total thyroidectomy is acceptable for tumors <4 cm in diameter. The decision to perform a total thyroidectomy may take into account disease on the contralateral side, the plan to administer postoperative radioiodine, and patient preference.

**Radioiodine therapy has not been shown to decrease recurrence rates in low-risk patients. The data on recurrence rates are equivocal in intermediate risk patients, and show benefit in high-risk patients.

***Anatomic recurrences typically involve disease in the cervical nodes. Treatment options include surgery or local ablative measures (eg, ethanol ablation), or observation for small (<1 cm) nodes, especially in the lateral compartments.

****Suppression of the serum TSH with levothyroxine has been shown to be of more benefit in patients with higher stages of disease. In older patients with thyroid cancer, the benefits of TSH suppression and iatrogenic thyrotoxicosis have to be balanced against a risk of atrial fibrillation and bone loss in postmenopausal women.

determined, and the patient is scanned 24 to 72 hours after a low dose of radioiodine (ie, 1-4 mCi of ^{131}I or 1-2 mCi ^{123}I). If there is evidence of residual RAIU in the neck or elsewhere radioactive iodine (^{131}I) is administered in a dose of 30 to 100 mCi, depending on the size and invasiveness of the primary tumor. A whole body radioiodine scan is then repeated 4 to 7 days later (called a **posttherapy scan**) to be sure no additional areas of RAIU are revealed. Nine to 12 months later, rhTSH is administered or thyroid hormone therapy is withdrawn, and a radioiodine scan and serum TG are repeated to document ablation of all functioning thyroid tissue. An undetectable serum TG level, at a time when the serum TSH is elevated, is the most sensitive evidence that all thyroid tissue has been eradicated. Indeed, a repeat whole body radioiodine scan may not be needed in low-risk patients. For the first year, patients are typically maintained on doses of T_4 that are adequate to suppress serum TSH to levels that are subnormal. Once it has been established that the patient is free of disease, the dose of T_4 can be increased to allow the serum TSH to rise into the low-normal range (0.3-2 mU/L).

rhTSH, rather than thyroid hormone withdrawal, is usually employed for follow-up evaluations. When using rhTSH, the patient continues to take his T_4, avoiding the adverse effects of hypothyroidism. If the patient's TSH-stimulated serum TG is less than 1 ng/mL and the whole body radioiodine scan, if obtained, is negative, the patient is likely free of disease. On the other hand, if the serum TG concentration rises above 2 ng/mL and/or if the radioiodine scan is positive, the patient is likely to have either persistent thyroid remnant tissue or residual thyroid cancer. Neck sonography, CT, MRI, or PET scanning should be considered in this circumstance to look for evidence of residual thyroid cancer. If disease is localized anatomically and proven to be thyroid cancer, additional surgery may be warranted. If the stimulated TG is greater than 20 ng/mL and other anatomic imaging studies are negative, some experts recommend empiric radioiodine therapy because of the high suspicion of residual disease that might then be visualized on the posttreatment scan.

Follow-up at intervals of 6 to 12 months should include careful examination of the neck, sometimes including neck ultrasound,

TABLE 7-16 Tumor (T), lymph node (N), and distant metastasis (M) classification and staging of thyroid cancer.

Definition of Primary Tumor (T)

Papillary, Follicular, and Hurthle Cell Thyroid Carcinoma

T Category	T Criteria
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Tumor ≤2 cm in greatest dimension limited to the thyroid
T1a	Tumor ≤1 cm in greatest dimension limited to the thyroid
T1b	Tumor >1 cm but ≤2 cm in greatest dimension limited to the thyroid
T2	Tumor >2 cm but ≤4 cm in greatest dimension limited to the thyroid
T3	Tumor >4 cm limited to the thyroid, or gross extrathyroidal extension invading only strap muscles
T3a	Tumor >4 cm limited to the thyroid
T3b	Gross extrathyroidal extension invading only strap muscles (sternohyoid, sternothyroid, thyrohyoid, or omohyoid muscles) from a tumor of any size
T4	Includes gross extrathyroidal extension
T4a	Gross extrathyroidal extension invading subcutaneous soft tissues, larynx, trachea, esophagus, or recurrent laryngeal nerve from a tumor of any size
T4b	Gross extrathyroidal extension invading prevertebral fascia or encasing the carotid artery or mediastinal vessels from a tumor of any size

Note: All categories may be subdivided: (s) solitary tumor and (m) multifocal tumor (the largest tumor determines the classification).

Definition of Distant Metastasis (M)

M Category	M Criteria
M0	No distant metastasis
M1	Distant metastasis

Definition of Regional Lymph Node (N)

N Category	N Criteria
NX	Regional lymph nodes cannot be assessed
N0	No evidence of locoregional lymph node metastasis
N0a	One or more cytologically or histologically confirmed benign lymph nodes
N0b	No radiologic or clinical evidence of locoregional lymph node metastasis
N1	Metastasis to regional nodes
N1a	Metastasis to level VI or VII (pretracheal, paratracheal, or prelaryngeal/Delphian, or upper mediastinal) lymph nodes. This can be unilateral or bilateral disease
N1b	Metastasis to unilateral, bilateral, or contralateral lateral neck lymph nodes (levels I, II, III, IV, or V) or retropharyngeal lymph nodes

AJCC Prognostic Stage Groups

Differentiated

When age at Diagnosis is...	And T is...	And N is...	And M is...	Then the Stage Group is...
<55 years	Any T	Any N	M0	I
<55 years	Any T	Any N	M1	II
≥55 years	T1	N0/NX	M0	I
≥55 years	T1	N1	M0	II
≥55 years	T2	N0/NX	M0	I
≥55 years	T2	N1	M0	II
≥55 years	T3a/T3b	Any N	M0	II
≥55 years	T4a	Any N	M0	III
≥55 years	T4b	Any N	M0	IVA
≥55 years	Any T	Any N	M1	IVB

Reproduced with permission from Amin MB, Edge SB, Greene FL, et al. *AJCC Cancer Staging Manual*, 8th ed. New York: Springer; 2017.

looking for recurrent masses, which typically develop in ipsilateral cervical nodes. If an abnormal lymph node is discovered, FNAB is indicated to confirm or rule out cancer depending on the size and location of the abnormal node(s). The patient's serum TSH should be checked to be certain it is adequately suppressed. The serum TG should be periodically assessed to be certain it is undetectable. A rise in serum TG to detectable levels while TSH is suppressed suggests tumor recurrence, and imaging studies such as neck ultrasound, CT, and MRI are necessary. PET scanning may also be useful to localize residual disease, especially in patients with very high serum TG levels (>20 ng/mL) when conventional imaging studies have been unrevealing.

TG autoantibodies present in 20% of patients interfere with accurate measurement of TG in radioimmunometric assays which are used in the majority of commercial laboratories and hospitals, causing falsely low values. Patients with TG autoantibodies must be followed with periodic imaging studies, such as thyroid ultrasound or CT scan.

The patient with a rising TG and a negative ^{131}I scan presents a common and difficult problem. Serial anatomic imaging, especially with cervical sonography, can sometimes identify the site of residual cervical disease. Administration of large empiric ^{131}I doses has not been shown to be of significant benefit. For patients with noniodine-avid metastatic disease that is progressive, recent studies have shown that several kinase–inhibiting drugs, including sorafenib and lenvatinib, can be effective in halting further tumor growth for 1 to 2 years. The kinases involved include the vascular endothelial growth factor receptor (VEGF) family, RET, and

TABLE 7–17 ATA system for estimating risk of thyroid cancer recurrence.*

LOW RISK (1-10% recurrence rate)
Papillary thyroid cancer with all of the following present:
- No local or distant metastases
- All macroscopic tumor has been resected
- No invasion of locoregional tissues
- Tumor does not have aggressive histology (eg, tall cell, insular, columnar cell carcinoma, Hürthle cell carcinoma, follicular thyroid cancer, hobnail variant)
- No vascular invasion
- No radioiodine uptake outside the thyroid bed on the post-treatment scan, if done
- Clinical N0 or ≤5 pathologic N1 micrometastases (<0.2 cm in largest dimension)
- Intrathyroidal, encapsulated follicular variant of papillary thyroid cancer*
- Intrathyroidal, well-differentiated follicular thyroid cancer with capsular invasion and no or minimal (<4 foci) vascular invasion
- Intrathyroidal, papillary microcarcinoma, unifocal or multifocal, including BRAF V600E mutated (if known)

INTERMEDIATE RISK (10-40% recurrence rate)
- Microscopic invasion into the perithyroidal soft tissues
- Cervical lymph node metastases or ^{131}I uptake outside the thyroid bed on the post-treatment scan done after thyroid remnant ablation
- Tumor with aggressive histology or vascular invasion (eg, tall cell, insular, columnar cell carcinoma, Hürthle cell carcinoma, follicular thyroid cancer, hobnail variant)
- Clinical N1 or >5 pathologic N1 with all involved lymph nodes <3 cm in largest dimension*
- Multifocal papillary thyroid microcarcinoma with extrathyroidal extension and BRAF V600E mutated (if known)*

HIGH RISK (>40% recurrence rate)
- Macroscopic tumor invasion
- Incomplete tumor resection with gross residual disease
- Distant metastases
- Postoperative serum thyroglobulin suggestive of distant metastases
- Pathologic N1 with any metastatic lymph node ≥3 in largest dimension*
- Follicular thyroid cancer with extensive vascular invasion (>4 foci of vascular invasion)

Reproduced with permission from Haugen BR Md, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016 Jan;26(1):1-133.

BRAF. For patients with bone or brain metastases, combined external radiation and ^{131}I therapy may be effective.

B. Differentiated thyroid cancer: course and prognosis

An optimal outcome is dependent on adequate therapy. There has been controversy over the extent of initial surgery for papillary and follicular thyroid cancer. As noted earlier, lesions under 4 cm with no evidence of local or distant metastases (T1, N0, M0) can probably be treated with lobectomy alone. However, in all other groups, total thyroidectomy and modified regional neck dissection (if gross evidence of spread is noted at the time of surgery) is indicated for two reasons: (1) it removes all local disease, and (2) it sets the stage for ^{131}I therapy and follow-up utilizing serum TG measurements. Total or near-total thyroidectomy must be performed by an experienced thyroid surgeon to minimize the complications of surgery.

A second factor in survival is the use of radioiodine for ablation of residual thyroid tissue after thyroidectomy in Stages 3 and 4 patients and the treatment of residual or recurrent disease. Low doses of 30 to 50 mCi ^{131}I are used to ablate residual thyroid tissue, but larger doses of 100 to 200 mCi are necessary for the treatment of invasive or metastatic disease. Acute adverse effects of the larger doses include radiation sickness, sialadenitis, gastritis, and transient oligospermia. Cumulative doses of ^{131}I above 500 mCi may be associated with transient infertility in females and azoospermia in males, pancytopenia in about 4% of patients, and leukemia in about 0.3%. Radiation pneumonitis may occur in patients with diffuse pulmonary metastases, but this is minimized by utilization of high-dose treatment no more than once a year. There is no evidence that radioiodine therapy improves the outcome in younger patients with low-risk disease, but it may decrease recurrence rates and overall survival in older patients and those with higher risk disease.

A third factor in survival is the adequate use of TSH suppression therapy. T_4 in a dose of 2.2 μg/kg/d (1 μg/lb/d) usually suppresses TSH to 0.1 mU/L or less, which removes a major growth factor for papillary or follicular thyroid cancer. Low-risk patients do not require suppressive doses of thyroid hormone, but larger doses are indicated in intermediate and higher risk patients. High-dose T_4 therapy is not without risk: there may be angina, tachycardia, or heart failure in older patients or tachycardia and nervousness in younger patients. In addition, there is an increased risk of osteoporosis in postmenopausal women. Calcium supplementation and bisphosphonate therapy may prevent bone loss in these patients, but the treatment program must be individualized.

C. Medullary carcinoma

Early and adequate initial thyroidectomy and cervical node dissection is the best therapy for medullary cancer. Once the disease has metastasized, it is very difficult to control, though the more favorable tumor types often progress very slowly. Patients with medullary cancer should be followed postoperatively with periodic measurement of serum markers (eg, calcitonin and CEA) that indicate residual disease. Family members of patients with an *RET* oncogene mutation should be screened for the mutation as noted earlier (see Figure 7–51). If a patient has a persistently elevated serum calcitonin concentration after total thyroidectomy and regional node dissection, neck ultrasound, CT, MRI, and/or selective venous catheterization and sampling for serum calcitonin may reveal the location of the metastases. Metastatic foci of medullary cancer may also be revealed by PET, indium-labeled somatostatin (octreotide), or sestamibi scans. If these studies fail to localize the lesion (as is often the case), the patient must be followed until the metastatic lesions become evident by physical examination or imaging studies. External x-ray therapy may be useful in the treatment of some metastatic lesions. Chemotherapy with tyrosine kinase inhibitors, especially the *RET* inhibitor vandetanib has recently been shown to be highly effective in

preventing disease progression in many patients with medullary cancer. Metastatic medullary cancer cannot be treated with ^{131}I, and TSH-suppressive doses of T_4 are not required.

Medullary thyroid cancer is most aggressive in patients with MEN 2B, less aggressive in the sporadic type, and least aggressive in MEN 2A and familial medullary thyroid cancer.

D. Anaplastic carcinoma   Anaplastic carcinoma of the thyroid has a very poor prognosis. Treatment consists of minimal surgery to biopsy the tumor and establish the diagnosis, and sometimes to debulk the tumor and prevent tracheal compression. Standard treatment incorporates combined radiotherapy and chemotherapy with doxorubicin, which is sometimes combined with other agents. Occasionally, patients have prolonged survival, but most succumb to their disease within 1 year.

REFERENCES

Physiology

Bernal J, Guadaño-Ferraz A, Morte B. Thyroid hormone transporters—functions and clinical implications. *Nat Rev Endocrinol*. 2015;11:406. [PMID: 25942657]

Bizhanova A, Kopp P. Minireview: the sodium-iodide symporter NIS and pendrin in iodide homeostasis of the thyroid. *Endocrinology*. 2009;150:1084. [PMID: 19196800]

Hammes SR, Davis PJ. Overlapping nongenomic and genomic actions of thyroid hormone and steroids. *Best Pract Res Clin Endocrinol Metab*. 2015;29:581. [PMID: 26303085]

Lonard DM, O'Malley BW. Nuclear receptor coregulators: modulators of pathology and therapeutic targets. *Nat Rev Endocrinol*. 2012;8:598. [PMID: 22733267]

McAninch EA, Bianco AC. Thyroid hormone signaling in energy homeostasis and energy metabolism. *Ann N Y Acad Sci*. 2014;1311:77. [PMID: 24697152]

Pappa T, Ferrara AM, Refetoff S. Inherited defects of thyroxine-binding proteins. *Best Pract Res Clin Endocrinol Metab*. 2015;29:735. [PMID: 26522458]

Sinha R, Yen PM. Cellular action of thyroid hormone. In: De Groot LJ, Beck-Peccoz P, Chrousos G, Dungan K, Grossman A, Hershman JM, Koch C, McLachlan R, New M, Rebar R, Singer F, Vinik A, Weickert MO, eds. *Endotext [Internet]*. South Dartmouth (MA): MDText.com, Inc.; 2000-2014 Feb 12.

Szkudlinski MW, Fremont V, Ronin C, Weintraub BD. Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure-function relationships. *Physiol Rev*. 2002;82:473. [PMID: 11917095]

Zimmermann MB. Iodine deficiency. *Endocr Rev*. 2009;30:376. [PMID: 19460960]

Tests of Thyroid Function

Fliers E, Bianco AC, Langouche L, Boelen A. Thyroid function in critically ill patients. *Lancet Diabetes Endocrinol*. 2015;3:816. [PMID: 26071885]

Hollowell JG, Staehling NW, Flanders WD, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*. 2002;87:489. [PMID: 11836274]

Koulouri O, Moran C, Halsall D, Chatterjee K, Gurnell M. Pitfalls in the measurement and interpretation of thyroid function tests. *Best Pract Res Clin Endocrinol Metab*. 2013;27:745. [PMID: 24275187]

Hypothyroidism

Biondi B, Wartofsky L. Treatment with thyroid hormone. *Endocr Rev*. 2014;35:433. [PMID: 24433025]

Cooper DS, Biondi B. Subclinical thyroid disease. *Lancet*. 2012;379:1142. [PMID: 22273398]

Grasberger H, Refetoff S. Genetic causes of congenital hypothyroidism due to dyshormonogenesis. *Curr Opin Pediatr*. 2011;23:421. [PMID: 21543982]

Jonklaas J, Bianco AC, Bauer AJ, et al. American Thyroid Association Task Force on Thyroid Hormone Replacement. Guidelines for the treatment of hypothyroidism: prepared by the American Thyroid Association task force on thyroid hormone replacement. *Thyroid*. 2014;24:1670. [PMID: 25266247]

Lazarus JH, Bestwick JP, Channon S, et al. Antenatal thyroid screening and childhood cognitive function. *N Engl J Med*. 2012;366:493. [PMID: 22316443]

Li M, Eastman CJ. The changing epidemiology of iodine deficiency. *Nat Rev Endocrinol*. 2012;8:434. [PMID: 22473332]

McDermott MT. Hypothyroidism. *Ann Intern Med*. 2009;151:61. [PMID: 19949140]

Negro R, Stagnaro-Green A. Diagnosis and management of subclinical hypothyroidism in pregnancy. *BMJ*. 2014;349:g4929. [PMID: 25288580]

Popoveniuc G, Chandra T, Sud A, et al. A diagnostic scoring system for myxedema coma. *Endocr Pract*. 2014;20:808. [PMID: 24518183]

Razvi S, Weaver JU, Butler TJ, Pearce SH. Levothyroxine treatment of subclinical hypothyroidism, fatal and nonfatal cardiovascular events, and mortality. *Arch Intern Med*. 2012;172:811. [PMID: 22529180]

Rodondi N, den Elzen WP, Bauer DC, et al. Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA*. 2010;304:1365. [PMID: 20858880]

Teng W, Shan Z, Patil-Sisodia K, Cooper DS. Hypothyroidism in pregnancy. *Lancet Diabetes Endocrinol*. 2013;1:228. [PMID: 24622371]

Hyperthyroidism

Bahn RS. Graves' ophthalmopathy. *N Engl J Med*. 2010;362:726. [PMID: 20181974]

Bahn RS, Burch HB, Cooper DS, et al. Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. *Thyroid*. 2011;21:593. [PMID: 21510801]

Bartalena L, Fatourechi V. Extrathyroidal manifestations of Graves' disease: a 2014 update. *J Endocrinol Invest*. 2014;37:691. [PMID: 24913238]

Burch HB, Cooper DS. Management of Graves disease: a review. *JAMA*. 2015;314:2544. [PMID: 26670972]

Cooper DS. Antithyroid drugs. *N Engl J Med*. 2005;352:905. [PMID: 15745981]

Cooper DS, Laurberg P. Hyperthyroidism in pregnancy. *Lancet Diabetes Endocrinol*. 2013;1:238. [PMID: 24622372]

De Leo S, Lee SY, Braverman LE. Hyperthyroidism. *Lancet*. 2016;388(10047):906. [PMID: 27038492]

Franklyn JA, Boelaert K. Thyrotoxicosis. *Lancet*. 2012;379:1155. [PMID: 22394559]

Iyer S, Bahn R. Immunopathogenesis of Graves' ophthalmopathy: the role of the TSH receptor. *Best Pract Res Clin Endocrinol Metab*. 2012;26:281. [PMID: 22632365]

Kraiem Z, Newfield RS. Graves' disease in childhood. *J Pediatr Endocrinol Metab*. 2001;229:43. [PMID: 11308041]

Laurberg P, Wallin G, Tallstedt L, Abraham-Nordling M, Lundell G, Torring O. TSH-receptor autoimmunity in Graves' disease after therapy with anti-thyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol*. 2008;158:69. [PMID: 18166819]

McDermott MT. Hyperthyroidism. *Ann Intern Med*. 2012;157:ITC1-16. [PMID: 22751778]

Nabil N, Miner DJ, Amatruda JM. Methimazole: an alternative route of administration. *J Clin Endocrinol Metab*. 1982;54:180. [PMID: 7054215]

Rhee CM, Bhan I, Alexander EK, Brunelli SM. Association between iodinated contrast media exposure and incident hyperthyroidism and hypothyroidism. *Arch Intern Med*. 2012;172:153. [PMID: 22271121]

Rivkees SA. Pediatric Graves' disease: controversies in management. *Horm Res Paediatr*. 2010;74:305. [PMID: 20924158]

Ross DS. Radioiodine therapy for hyperthyroidism. *N Engl J Med*. 2011;364:542. [PMID: 21306240]

Stein JD, Childers D, Gupta S, et al. Risk factors for developing thyroid-associated ophthalmopathy among individuals with Graves disease. *JAMA Ophthalmol.* 2015;133:290. [PMID: 25502604]

Sun S, Qiu X, Zhou J. Clinical analysis of 65 cases of hyperemesis gravidarum with gestational transient thyrotoxicosis. *J Obstet Gynaecol Res.* 2014;40:1567. [PMID: 24888917]

Amiodarone Related Thyroid Disease

Bogazzi F, Tomisti L, Bartalena L, Aghini-Lombardi F, Martino E. Amiodarone and the thyroid: a 2012 update. *J Endocrinol Invest.* 2012;35:340. [PMID: 22433945]

Thyroid Hormone Resistance

Ortiga-Carvalho TM, Sidhaye AR, Wondisford FE. Thyroid hormone receptors and resistance to thyroid hormone disorders. *Nat Rev Endocrinol.* 2014;10:582. [PMID: 25135573]

Syndrome of Inappropriate TSH Secretion

Beck-Peccoz P, Lania A, Beckers A, Chatterjee K, Wemeau JL. 2013 European thyroid association guidelines for the diagnosis and treatment of thyrotropin-secreting pituitary tumors. *Eur Thyroid J.* 2013;2:76. [PMID: 24783044]

Multinodular Goiter

Hashmi SM, Premachandra DJ, Bennett AM, Parry W. Management of retrosternal goitres: results of early surgical intervention to prevent airway morbidity, and a review of the English literature. *J Laryngol Otol.* 2006;120:644. [PMID: 16884549]

Knobel M. Etiopathology, clinical features, and treatment of diffuse and multinodular nontoxic goiters. *J Endocrinol Invest.* 2016;39:357. [PMID: 26392367]

Paschke R. Molecular pathogenesis of nodular goiter. *Langenbecks Arch Surg.* 2011;396:1127. [PMID: 21487943]

Thyroiditis

Fatourechi V, Aniszewski JP, Fatourechi GZ, Atkinson EJ, Jacobsen SJ. Clinical features and outcome of subacute thyroiditis in an incidence cohort: Olmsted County, Minnesota. *J Clin Endocrinol Metab.* 2003;88:2100. [PMID: 12727961]

Samuels MH. Subacute, silent, and postpartum thyroiditis. *Med Clin North Am.* 2012;96:223. [PMID: 22443972]

Stagnaro-Green A. Approach to the patient with postpartum thyroiditis. *J Clin Endocrinol Metab.* 2012;97:334. [PMID: 22312089]

Tomer Y. Mechanisms of autoimmune thyroid diseases: from genetics to epigenetics. *Annu Rev Pathol.* 2014;9:147. [PMID: 24460189]

Radiation Exposure

Inskip PD. Thyroid cancer after radiotherapy for childhood cancer. *Med Pediatr Oncol.* 2001;36:568. [PMID: 11340614]

Rubino C, Cailleux AF, De Vathaire F, Schlumberger M. Thyroid cancer after radiation exposure. *Eur J Cancer.* 2002;38:645. [PMID: 11916545]

Thyroid Nodules and Thyroid Cancer

Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med.* 2012;367:705. [PMID: 22731672]

Brito JP, Gionfriddo MR, Al Nofal A, et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. *J Clin Endocrinol Metab.* 2014;99:1253. [PMID: 24276450]

Brito JP, Hay ID, Morris JC. Low risk papillary thyroid cancer. *BMJ.* 2014;348:g3045. [PMID: 24935445]

Burman KD, Wartofsky L. Thyroid nodules. *N Engl J Med.* 2015;373:2347. [PMID: 26650154]

Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. *Thyroid.* 2009;19:1159. [PMID: 19846805]

Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg.* 2014;140:317. [PMID: 24557566]

Durante C, Costante G, Lucisano G, et al. The natural history of benign thyroid nodules. *JAMA.* 2015;313:926. [PMID: 25734734]

Francis GL, Waguespack SG, Bauer AJ, et al. Management Guidelines for Children with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid.* 2015;25:716. [PMID: 25900731]

Haddad R, Sherman SI, Shah JP, Wirth LJ. New frontiers and treatment paradigms for thyroid carcinoma. *Clin Adv Hematol Oncol.* 2014;12(7 Suppl 14):3. [PMID: 25768715]

Haugen BR Md, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid.* 2016;26:1. [PMID: 26462967]

Mallick U, Harmer C, Yap B, et al. Ablation with low-dose radioiodine and thyrotropin alfa in thyroid cancer. *N Engl J Med.* 2012;366:1674. [PMID: 22551128]

McLeod DS, Sawka AM, Cooper DS. Controversies in primary treatment of low-risk papillary thyroid cancer. *Lancet.* 2013;381:1046. [PMID: 23668555]

Moon WJ, Jung SL, Lee JH, et al. Benign and malignant thyroid nodules: US differentiation—multicenter retrospective study. *Radiology.* 2008;247:762. [PMID: 18403624]

Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. *Nat Rev Endocrinol.* 2011;7:569. [PMID: 21878896]

Smallridge RC, Ain KB, Asa SL, et al. American Thyroid Association American Thyroid Association guidelines for management of patients with anaplastic thyroid cancer. *Thyroid.* 2012;22:1104. [PMID: 23130564]

Wells SA Jr, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid.* 2015;25:567. [PMID: 25810047]

Xing M, Alzahrani AS, Carson KA, et al. Association between BRAF V600E mutation and mortality in patients with papillary thyroid cancer. *JAMA.* 2013;309:1493. [PMID: 23571588]

This page intentionally left blank

Metabolic Bone Disease

Dolores M. Shoback, MD, Anne L. Schafer, MD, and Daniel D. Bikle, MD, PhD

ACTH	Adrenocorticotropin hormone		**NALP5**	NACHT leucine-rich-repeat protein 5
ADHR	Autosomal dominant hypophosphatemic rickets		**OPG**	Osteoprotegerin
AHO	Albright hereditary osteodystrophy		**PHP**	Pseudohypoparathyroidism
AIRE	Autoimmune regulator		**PIP$_2$**	Phosphatidylinositol 4,5-bisphosphate
BMD	Bone mineral density		**PPHP**	Pseudopseudohypoparathyroidism
CaSR	Extracellular calcium-sensing receptor		**PTH**	Parathyroid hormone
CGRP	Calcitonin gene–related peptide		**PTHrP**	Parathyroid hormone–related protein
DBP	Vitamin D–binding protein		**RANK**	Receptor activator of nuclear factor kappa B
DXA	Dual-energy x-ray absorptiometry		**RANKL**	Receptor activator of nuclear factor kappa-B ligand
FGF23	Fibroblast-derived growth factor 23		**RAR**	Retinoic acid receptor
FHH	Familial hypocalciuric hypercalcemia		**RBP**	Retinol-binding protein
GALNT3	UDP-*N*-acetyl-α-D-galactosamine transferase		**RET**	Rearranged during transfection
HPT-JT	Hyperparathyroidism-jaw tumor		**RXR**	Retinoid X receptor
HT	Hormone therapy		**SERMs**	Selective estrogen response modulators
ICMA	Immunochemiluminescent assay		**sFRP**	Secreted frizzled related protein
IFN	Interferon		**TNF**	Tumor necrosis factor
IGF	Insulin-like growth factor		**TPN**	Total parenteral nutrition
IL	Interleukin		**TRP**	Tubular reabsorption of phosphate
IP$_3$	Inositol 1,4,5-triphosphate		**VDR**	Vitamin D receptor
IRMA	Immunoradiometric assay		**VDRE**	Vitamin D response element
LDL	Low-density lipoprotein		**VIP**	Vasoactive intestinal polypeptide
MCT	Medullary carcinoma of thyroid		**WHI**	Women's Health Initiative
MEN	Multiple endocrine neoplasia		**WHO**	World Health Organization
MEPE	Matrix extracellular phosphoglycoprotein		**XLH**	X-linked hypophosphatemia

CELLULAR AND EXTRACELLULAR CALCIUM METABOLISM

The calcium ion plays a critical role in intracellular and extracellular events in human physiology. Extracellular calcium levels in humans are tightly regulated within a narrow physiologic range to provide for proper functioning of many tissues: excitation-contraction coupling in the heart and other muscles, synaptic transmission and other functions of the nervous system, platelet aggregation, coagulation, and secretion of hormones and other regulators by exocytosis. The level of intracellular calcium is also tightly controlled, at levels about 10,000-fold lower than extracellular calcium, in order for calcium to serve as an intracellular

TABLE 8–1 Calcium concentrations in body fluids.

Total serum calcium	8.5-10.5 mg/dL (2.1-2.6 mmol/L)
Ionized calcium	4.4-5.2 mg/dL (1.1-1.3 mmol/L)
Protein-bound calcium	4.0-4.6 mg/dL (0.9-1.1 mmol/L)
Complexed calcium	0.7 mg/dL (0.18 mmol/L)
Intracellular free calcium	0.00018 mmol/L (180 nmol/L)

second messenger in the regulation of cell division, muscle contractility, cell motility, membrane trafficking, and secretion.

It is the concentration of ionized calcium ($[Ca^{2+}]$) that is regulated in the extracellular fluid. The ionized calcium concentration averages 1.25 ± 0.07 mmol/L (Table 8–1). However, only about 50% of the total calcium in serum and other extracellular fluids is present in the ionized form. The remainder is bound to albumin (about 40%) or complexed with anions such as phosphate and citrate (about 10%). The protein-bound and complexed fractions of serum calcium are metabolically inert and are not regulated by hormones; only the ionized $[Ca^{2+}]$ serves a regulatory role, and only this fraction is itself regulated by the calciotropic hormones parathyroid hormone (PTH) and 1,25 dihydroxyvitamin D [$1,25(OH)_2D$]. Large increases in the serum concentrations of phosphate or citrate can, however, by mass action, markedly increase the complexed fraction of calcium. For example, massive transfusions of whole blood, in which citrate is used as an anticoagulant, can reduce the ionized $[Ca^{2+}]$ enough to produce tetany. In addition, because calcium and phosphate circulate at concentrations close to saturation, a substantial rise in the serum concentration of either calcium or phosphate can lead to the precipitation of calcium phosphate salts in tissues. This is a source of major clinical problems in patients with severe hypercalcemia (eg, malignant tumors) and in those with severe hyperphosphatemia (eg, in renal failure or rhabdomyolysis).

What is remarkable about calcium metabolism is that the serum-ionized $[Ca^{2+}]$, which represents a tiny fraction of the total body calcium, can be so tightly regulated in the face of the rapid fluxes of calcium through body compartments that take place during the course of calcium metabolism (Figure 8–1). The total calcium in extracellular fluid amounts to about 1% of total body calcium, with most of the remainder sequestered in bone. Yet from the extracellular fluid compartment, which contains about 900 mg of calcium, 10,000 mg/d is filtered at the glomerulus and 500 mg/d is added to a labile pool in bone; and to the extracellular fluid compartment are added about 200 mg absorbed from the diet, 9800 mg reabsorbed by the renal tubule, and 500 mg from bone.

The challenge of calcium homeostasis, then, is to maintain a constant level of ionized $[Ca^{2+}]$ in the extracellular fluid, simultaneously providing adequate amounts of calcium to cells, to bone, and for renal excretion—and all the while compensating, on an hourly basis, for changes in daily intake of calcium, bone metabolism, and renal function. It is scarcely surprising that this homeostatic task requires two hormones, PTH and $1,25(OH)_2D$, or that the secretion of each hormone is exquisitely sensitive to small changes in the serum calcium, or that each hormone is able to regulate calcium exchange across all three interfaces of the extracellular fluid: the gut, the bone, and the renal tubule. This chapter reexamines the integrated roles of PTH and $1,25(OH)_2D$ in calcium homeostasis after their actions and secretory control have been described.

The challenge of the cellular calcium economy is to maintain an intracellular $[Ca^{2+}]$ or $[Ca^{2+}]_i$ of ~100 nmol/L, about 10,000-fold less than what is present outside cells (~1.0 mmol/L), providing for rapid fluxes through the intracellular compartment, as required for regulation, while maintaining a large gradient across the cell membrane. The calcium gradient across the cell membrane is maintained by ATP-dependent calcium pumps, Na^+-Ca^{2+} exchangers, and the storage of calcium within intracellular sites. Calcium can enter cells through several types of Ca^{2+} channels, some of which are voltage-operated or receptor-operated, to provide for rapid influx in response to depolarization or receptor stimulation. The cell also maintains large stores of calcium in microsomal and mitochondrial pools and in some cells in the Golgi. Calcium can be released from microsomal stores rapidly by cellular signals such as 1,4,5-inositol trisphosphate (IP_3) (see Chapter 1). Reuptake mechanisms are also present, so that intracellular Ca^{2+} transients can be rapidly terminated by returning calcium to storage pools or pumping it across the plasma membrane.

PARATHYROID HORMONE

Anatomy and Embryology of the Parathyroid Glands

PTH is secreted from four glands located adjacent to the thyroid gland in the neck. The glands weigh an average of 40 mg each. The two superior glands are usually found near the posterior aspect of the thyroid capsule; the inferior glands are most often located near the inferior thyroid margin. However, the exact location of the glands is variable, and 12% to 15% of normal persons

FIGURE 8–1 Calcium fluxes in a normal individual in a state of zero external mineral balance. The blue arrows denote unidirectional calcium fluxes; the pink arrows denote net fluxes. (Reproduced with permission from Felig P, Baxter JD, Broadus AE, et al. *Endocrinology and Metabolism*. 2nd ed. New York: McGraw-Hill Education; 1987.)

have a fifth parathyroid gland. The parathyroid glands arise from the third and fourth branchial pouches. The inferior glands are actually those derived from the third branchial pouches. Beginning cephalad to the other pair, they migrate further caudad, and one of them sometimes follows the thymus gland into the superior mediastinum. The small size of the parathyroids and the vagaries of their location and number make parathyroid surgery a challenging procedure for all but the expert surgeon.

The parathyroid glands are composed of epithelial cells and stromal fat. The predominant epithelial cell is the **chief cell**. The chief cell is distinguished by its clear cytoplasm from the **oxyphil cell**, which is slightly larger and has eosinophilic granular cytoplasm. Both cell types contain PTH, and it is not known whether their secretory regulation differs in any fundamental way.

Secretion of Parathyroid Hormone

To regulate the extracellular calcium concentration ($[Ca^{2+}]_e$), PTH is under tight control by the serum $[Ca^{2+}]$. Thus, the negative feedback relationship of PTH with serum $[Ca^{2+}]$ is steeply sigmoidal, with the steep portion of the curve corresponding exactly to the normal range of serum $[Ca^{2+}]$—precisely the relationship to create a high "gain" controller and ensure maintenance of the normal serum ionized $[Ca^{2+}]$ by PTH (Figure 8–2).

To sense the serum ionized $[Ca^{2+}]$ and thereby regulate the secretion of PTH, the parathyroid cell relies on relatively high levels of expression of the extracellular calcium-sensing receptor (CaSR). CaSR is a 120-kDa G protein–coupled receptor belonging to family C of this superfamily (see Chapter 1). The CaSR has

sequence homologies to the metabotropic glutamate receptors of the central nervous system, the type B gamma-aminobutyric acid receptor, and a large family of pheromone receptors. The large extracellular domain of the CaSR mediates the sensing of calcium and other ions. Like other G protein–coupled receptors, the CaSR has seven membrane-spanning domains. The intracellular loops that connect these domains are directly involved in coupling the receptor to G proteins.

Mutations in this receptor are responsible for one form of familial hypocalciuric hypercalcemia (FHH), a disorder of calcium-sensing by the parathyroid glands and kidney. The CaSR is not unique to the parathyroid. CaSRs are widely distributed in the brain, skin, growth plate cartilage, bone, intestine, stomach, C cells, and other tissues. This receptor regulates the responses to calcium in thyroid C cells, which secrete calcitonin in response to high $[Ca^{2+}]_e$, and in the distal nephron of the kidney, where the receptor regulates both calcium excretion and water clearance among other important functions. The role of CaSRs in many other tissues is still unclear.

The primary intracellular signal by which an increased $[Ca^{2+}]_e$ inhibits the secretion of PTH appears to be an increase in $[Ca^{2+}]_i$. The CaSR is directly coupled through G_q to the enzyme phospholipase C, which hydrolyzes the phospholipid phosphatidylinositol 4,5-bisphosphate (PIP_2) to liberate the intracellular messengers IP_3 and diacylglycerol (see Chapter 1). IP_3 binds to a receptor in endoplasmic reticulum that releases calcium from membrane stores. The release of stored calcium raises the $[Ca^{2+}]_i$ rapidly and is followed by a sustained influx of extracellular calcium, through channels that produce a rise and sustained plateau in $[Ca^{2+}]_i$. Increased $[Ca^{2+}]_i$ may be sufficient for inhibition of PTH release, but it is unclear whether calcium release from intracellular stores or sustained calcium influx from the cell exterior is the more important source for calcium. The other product of phospholipase C action is the lipid diacylglycerol, an activator of the calcium- and phospholipid-sensitive protein kinases in the protein kinase C family. The effects of protein kinase C isoenzymes on the release of PTH from the gland are complex. CaSRs also couple to the inhibition of cyclic adenosine-3′,5′-monophosphate (cAMP) generation, which also may play a role in setting the response of parathyroid cells to ambient calcium levels, as well as other signaling pathways.

The initial effect of high $[Ca^{2+}]_e$ is to inhibit the secretion of preformed PTH from storage granules in the gland by blocking the fusion of storage granules with the cell membrane and release of their contents. In most cells, stimulation of exocytosis (*stimulus-secretion coupling*) is a calcium-requiring process, which is inhibited by depletion of extracellular calcium. The parathyroid cell is necessarily an exception to this rule, because this cell must increase secretion of PTH when the ionized $[Ca^{2+}]$ is low. In the parathyroids, intracellular magnesium may serve the role in stimulus-secretion coupling that calcium does in other cells. As discussed later in the section on hypoparathyroidism, depletion of magnesium stores can paralyze the secretion of PTH, leading to reversible hypoparathyroidism.

Besides calcium, there are several regulators of PTH secretion. Hypermagnesemia inhibits PTH, and during the treatment of
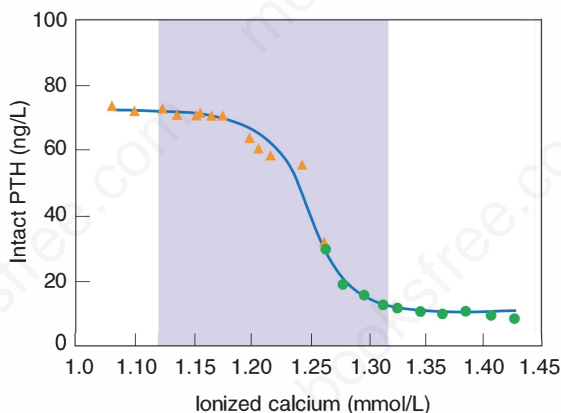
FIGURE 8–2 The relationship between the serum ionized $[Ca^{2+}]$ and the simultaneous serum concentration of intact PTH in normal humans. The serum calcium concentration was altered by the infusion of calcium (closed circles) or citrate (closed triangles). Parathyroid sensitivity to changes in serum calcium is maximal within the normal range (the shaded area). Low concentrations of PTH persist in the face of hypercalcemia. (Modified with permission from Conlin PR, et al. Hysteresis in the relationship between serum ionized calcium and intact parathyroid hormone during recovery from induced hyper- and hypocalcemia in normal humans. *J Clin Endocrinol Metab.* 1989 Sep;69(3):593-599.)

premature labor with infusions of magnesium sulfate, reductions in PTH levels and occasionally hypocalcemia are observed. Conversely, moderate hypomagnesemia can stimulate PTH secretion, even though prolonged depletion of magnesium will paralyze it. On a molar basis, magnesium is less potent in controlling secretion than calcium. Catecholamines, acting through β-adrenergic receptors and cAMP, stimulate the secretion of PTH. This effect does not appear to be clinically significant. The hypercalcemia sometimes observed in patients with pheochromocytoma usually has another basis—secretion of parathyroid hormone–related protein (PTHrP) by the tumor.

Not only do changes in serum ionized $[Ca^{2+}]$ regulate the secretion of PTH—they also regulate the synthesis of PTH at the level of stabilizing preproPTH mRNA levels. It is estimated that glandular stores of PTH are sufficient to maintain maximal rates of secretion for no more than 1.5 hours, so increased synthesis is required to meet sustained hypocalcemic challenges.

Transcription of the *PTH* gene is also regulated by vitamin D metabolites: high levels of 1,25(OH)₂D inhibit *PTH* gene transcription. This is one of many ways that the calciotropic hormones cooperatively regulate calcium homeostasis, and it has therapeutic implications. Vitamin D analogs are used to treat secondary hyperparathyroidism in dialysis patients.

Synthesis and Processing of Parathyroid Hormone

PTH is an 84-amino-acid peptide with a molecular weight of 9300. Its gene is located on chromosome 11. The gene encodes a precursor called preproPTH with a 29-amino-acid extension added at the amino terminus of the mature PTH peptide (Figure 8–3). This extension includes a 23-amino-acid signal sequence (the *pre* sequence) and a 6-residue prohormone sequence. The signal sequence in preproPTH functions precisely as it does in most other secreted protein molecules, to allow recognition of the peptide by a signal recognition particle, which binds to nascent peptide chains as they emerge from the ribosome and guides them to the endoplasmic reticulum, where they are inserted through the membrane into the lumen (Figure 8–4).

In the lumen of the endoplasmic reticulum, a signal peptidase cleaves the signal sequence from preproPTH to leave proPTH, which exits the endoplasmic reticulum and travels to the Golgi, where the *pro* sequence is cleaved from PTH by an enzyme called furin. Whereas preproPTH is evanescent, proPTH has a life span of about 15 minutes. The processing of proPTH is quite efficient, and proPTH, unlike other prohormones (eg, proinsulin), is not secreted. As it leaves the Golgi, PTH is repackaged into dense neuroendocrine-type secretory granules, where it is stored to await secretion.

FIGURE 8–3 Primary structure of human preproPTH hormone. The arrows indicate sites of specific cleavages which occur in the sequence of biosynthesis and peripheral metabolism of the hormone. The biologically active sequence is enclosed in the center of the molecule. (Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism*, 3rd ed. New York: McGraw-Hill Education; 1995.)

FIGURE 8–4   Biosynthetic events in the production of PTH within the parathyroid cell. *PreproPTH* gene is transcribed to its mRNA, which is translated on the ribosomes to preproPTH (amino acids –29 to +84). The presequence is removed within the endoplasmic reticulum, yielding proPTH (–6 to +84). Mature PTH (1-84) released from the Golgi is packaged in secretory granules and released into the circulation –in the presence of hypocalcemia. The CaSR, or CaR, senses changes in extracellular calcium that affect both the release of PTH and the transcription of the *preproPTH* gene. (Reproduced with permission from McPhee SJ and Lingappa VR. *Pathophysiology of Disease: An Introduction to Clinical Medicine.* New York: McGraw-Hill; 1995.)

Clearance and Metabolism of PTH

PTH secreted by the gland has a circulating half-life of 2 to 4 minutes. Intact PTH(1-84) is predominantly cleared in the liver and kidney. There, PTH is cleaved at the 33 to 34 and 36 to 37 positions to produce an amino terminal fragment and a carboxyl terminal fragment. Amino terminal fragments of PTH do not circulate to the same extent as carboxyl terminal fragments. The latter are cleared from blood by renal filtration, and they accumulate in chronic renal failure. Although the classic activities of PTH are encoded in the amino terminal portion of the molecule, mid region and carboxyl terminal fragments of the hormone may not be metabolically inert. These fragments may have their own receptors and biologic actions.

Assays of PTH

Current assays of intact PTH(1-84) employ two-site immunoradiometric assay (IRMA) or immunochemiluminescent assay (ICMA) techniques, in which the normal range for PTH is approximately 10 to 60 pg/mL (1-6 pmol/L). By utilizing antibodies to two determinants, one near the amino terminal end of the PTH molecule and the other near the carboxyl terminal end of PTH, these assays are designed to measure the intact, biologically active hormone species (Figure 8–5). In practice, such assays have sufficient sensitivity and specificity to detect not only increased levels of PTH in patients with hyperparathyroidism but also suppressed levels of PTH in patients with nonparathyroid hypercalcemia. The ability to detect suppression of PTH makes these assays powerful tools for the differential diagnosis of hypercalcemia. If hypercalcemia results from a form of hyperparathyroidism, then the serum PTH level will be high; if hypercalcemia has a *nonparathyroid* basis, then PTH will be suppressed.

Recently, intact PTH assays have been further refined so that the amino-terminal antibody recognizes only the most amino terminal residues of PTH (essentially amino acids 1-6). This improvement was made because it became clear that antibodies used in first-generation, two-site intact PTH assays were detecting species of PTH that were not PTH(1-84). There are
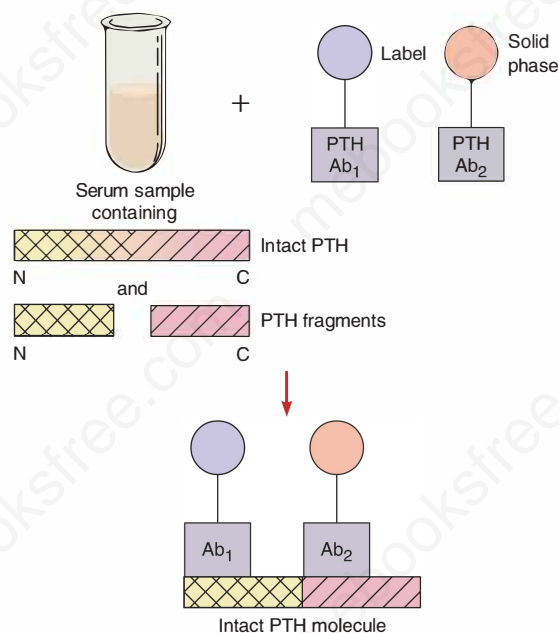
FIGURE 8–5 Schematic representation of the principle of the two-site assay for intact PTH. The label may be a luminescent probe or ^{125}I in the immunochemiluminescent or IRMA, respectively. Two different region-specific antibodies are used (Ab$_1$ and Ab$_2$). Only the hormone species containing both immunodeterminants is counted in the assay. Most recent whole or intact PTH immunoassays employ an Ab$_1$ directed against extreme *N*-terminal PTH epitopes as shown. (Reproduced with permission from McPhee SJ and Lingappa VR. *Pathophysiology of Disease: An Introduction to Clinical Medicine.* New York: McGraw-Hill; 1995.)

non-PTH(1-84) amino-terminally truncated PTH fragments generated *in vivo* by peripheral metabolism that can accumulate in patients with chronic renal failure [eg, PTH(7-84)]. About 50% of circulating PTH in patients with renal failure consists of large fragments of non-(1-84)PTH, and there is individual variation in these percentages. Assays specific for full-length PTH, in which the amino terminal antibodies truly detect residues at the extreme *N*-terminus of the molecule (residues 1-6), are designated as *whole, bioactive,* or *intact* PTH.

Biologic Effects of PTH

The central function of PTH is to regulate serum ionized [Ca^{2+}] levels by concerted effects on three principal target organs: bone, intestinal mucosa, and kidney. The effect of PTH on intestinal calcium absorption is indirect, resulting from increased production of the intestinally active vitamin D metabolite 1,25(OH)$_2$D. By its integrated effects on the kidney, gut, and bone, PTH acts to increase the inflow of calcium into the extracellular fluid and thus defend against hypocalcemia. Removal of the parathyroid glands results in profound hypocalcemia and ultimately in tetany and even death.

In the kidney, PTH has direct effects on the tubular reabsorption of calcium, phosphate, and bicarbonate. Although the bulk of calcium is resorbed from tubule fluid together with sodium in the proximal convoluted tubule, the fine-tuning of calcium excretion occurs in the distal nephron. There, PTH markedly increases the reabsorption of calcium, predominantly in the distal convoluted tubule. Although calcium is actively transported against an electrochemical gradient, the precise nature of the calcium transport process that is regulated by PTH is uncertain. However, from a physiologic standpoint, the ability to limit renal losses of calcium is one important means by which PTH defends the serum ionized [Ca^{2+}].

PTH inhibits the reabsorption of phosphate in the renal proximal tubule. In this nephron segment, phosphate is transported across the apical membrane of the tubule cell by specific sodium-phosphate cotransporters, with phosphate influx driven by the energy of the sodium gradient. PTH inhibits sodium-phosphate reabsorption by reducing the rate of insertion of transporters from a sequestered cytoplasmic pool into the apical membrane. The phosphaturic effect of PTH is profound. It is best quantified by calculating the tubular reabsorption of phosphate (TRP) from the clearances of phosphate and creatinine (TRP = $1 - C_p/C_{creat}$, normal range 80%-97%), or by calculating the renal phosphate threshold (TmP/GFR) from a standard nomogram. Because it is primarily the renal phosphate threshold that sets the level of serum phosphate, the phosphaturic effect of PTH is mirrored in the serum phosphorus level (eg, hypophosphatemia in hyperparathyroidism). Hyperparathyroid states may also be characterized by impaired bicarbonate reabsorption and a mild hyperchloremic metabolic acidosis, because of inhibition of Na$^+$-H$^+$ antiporter activity by PTH.

Although the hypocalciuric effect of PTH is readily understood as part of the concerted actions of the hormone to defend the serum ionized [Ca^{2+}], the utility of the phosphaturic effect of PTH is less obvious. One consideration is that the phosphaturic effect tends to prevent an increase in serum phosphate, which would otherwise result from the obligatory release of phosphate with calcium during bone resorption and would tend to dampen the homeostatic increase in serum ionized [Ca^{2+}] by complexing calcium in blood. An example is chronic kidney disease (CKD). When phosphate clearance is impaired by renal failure, the hypocalcemic effect of phosphate released during bone remodeling is an important contributor to progressive secondary hyperparathyroidism as part of a positive feedback loop—the more that bone resorption is stimulated and phosphate released, the more hyperparathyroidism is induced. Under those conditions, high serum levels of phosphate also stimulate PTH secretion.

Mechanism of Action of Parathyroid Hormone

There are two mammalian receptors for PTH. The first receptor to be identified recognizes PTH and PTHrP and is designated the PTH/PTHrP receptor, or the PTH-1 receptor (PTH-1R). The PTH-2R is activated by PTH only. These receptors also differ in their tissue distribution. The PTH-1R in kidney and bone is an 80,000-MW glycoprotein and member of the G protein receptor superfamily. It has the canonical architecture of such receptors,

with a large first extracellular domain, 7 membrane-spanning domains, and a cytoplasmic tail (see Chapter 1). PTH binds to sites in the large extracellular domain of the receptor. The hormone-bound form of the receptor then activates associated G proteins via several determinants in the intracellular loops. PTH receptors are closely related to a small subfamily of peptide hormone receptors, which includes those for secretin, vasoactive intestinal polypeptide (VIP), adrenocorticotropin hormone (ACTH), and calcitonin.

PTH itself has a close structural resemblance to a sister protein, PTHrP, and also resembles the peptides secretin, VIP, calcitonin, and ACTH. As noted earlier, the receptors for these related ligands are themselves members of a receptor subfamily. These peptide hormones have an amino terminal alpha helical domain, thought to be directly involved in receptor activation, and an adjacent alpha helical domain that seems to be the primary receptor-binding domain. In the case of PTH, residues 1 to 6 are required for activation of the receptor; truncated analogs without these residues [eg, PTH(7-34)] can bind the receptor but cannot efficiently activate it and thus serve as competitive antagonists of PTH action. The primary receptor-binding domain consists of PTH(18-34). Although the intact form of PTH is an 84-amino-acid peptide, PTH(35-84) does not seem to have any important role in binding to the bone-kidney receptor. However, a separate PTH(35-84) receptor may exist; the carboxyl terminal PTH receptor could mediate an entirely different set of actions of PTH.

The PTH-1 receptor binds PTH and PTHrP with equal affinity. Physiologic activation of the receptor by binding of either PTH or PTHrP induces the active, GTP-bound state of two receptor-associated G proteins. G_s couples the receptor to the effector adenylyl cyclase and thereby to the generation of cAMP as a cellular second-messenger. G_q couples the receptor to a separate effector system, phospholipase C, and thereby to an increase in $[Ca^{2+}]_i$ and to activation of protein kinase C (see Chapter 1). Although it is not clear which of the cellular messengers, cAMP, $[Ca^{2+}]_i$, or diacylglycerol is responsible for each of the various cellular effects of PTH, there is evidence from an experiment of nature that cAMP is the key second-messenger for maintaining calcium homeostasis and renal phosphate excretion. The experiment is pseudohypoparathyroidism (PHP), in which null mutations in one allele of the stimulatory G protein subunit G_s alpha (GNAS) cause hypocalcemia and unresponsiveness of renal phosphate excretion to PTH.

PTHrP

When secreted in abundance by malignant tumors, PTHrP produces severe hypercalcemia by activating the PTH-1 receptor. However, the physiologic roles of PTHrP are quite different from those of PTH. PTHrP is produced in many fetal and adult tissues. Based on gene knockout experiments and overexpression of PTHrP in individual tissues, we now know that PTHrP is required for normal development as a regulator of the proliferation and mineralization of chondrocytes and as a regulator of placental calcium transport. In postnatal life, PTHrP also appears to regulate epithelial-mesenchymal interactions that are critical for development of the mammary gland, skin, and hair follicle. In most physiologic circumstances, PTHrP carries out local rather than systemic actions. PTHrP is discussed fully in Chapter 21.

Calcitonin

Calcitonin (CT) is a 32-amino-acid peptide whose principal function is to inhibit osteoclast-mediated bone resorption. CT is secreted by parafollicular C cells of the thyroid. These are neuroendocrine cells derived from the ultimobranchial body, which fuses with the posterior lobes of the thyroid to become the C cell mass (about 0.1% of the thyroid gland).

The secretion of CT is under the control of the ionized $[Ca^{2+}]$. The C cell uses the same CaSR as the parathyroid cell to sense changes in the ambient ionized $[Ca^{2+}]$. In contrast to parathyroid cells, C cells increase secretion of CT in response to hypercalcemia and shut off hormone secretion during hypocalcemia.

CT is a member of the CT gene–related peptide (CGRP) superfamily that includes: CT, CGRP, CT-receptor–stimulating peptide (CRSP), adrenomedullin, and amylin. CT and CGRP are encoded by one gene, and the respective peptides are generated from alternative splicing of primary RNA transcripts in a tissue-selective manner, with CT made predominantly in the thyroid C cells. The genomic structure of the *CT/CGRP/CRSP* gene family is shown in Figure 8–6. In humans, there are two genes (*CALCA* and *CALCB*) that encode CGRPs and CT and one pseudogene (*CALCP*). Transcription of CALCA produces CT and CGRP in a tissue-specific manner by alternative splicing. The *CALCB* gene has a coding CGRP-beta exon and an exon that encodes for CT, but in which there is an in-frame stop codon. Thus, CALCB generates only CGRP-beta. In the thyroid C cell, mature CT (Figure 8–7) is incorporated within a
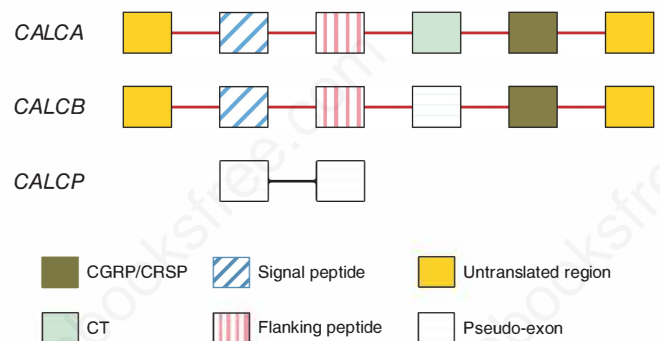
FIGURE 8–6 Genomic structure of the human calcitonin (CT)/ calcitonin gene–related peptide (CGRP)/CT-receptor–stimulating peptide (CRSP) family of genes which includes CALCA and CALCB and the pseudogene CALCP. Alternative splicing generates the two gene products (CT in the thyroid C cells and CGRP in the nervous system) with well-defined functions in humans. Straight lines indicate the introns while the coding sequences are shown by specifically designated boxes. (Data from Katafuchi T, Yasue H, Osaki T, et al. Calcitonin receptor-stimulating peptide: its evolutionary and functional relationship with calcitonin/calcitonin gene-related peptide based on gene structure. *Peptides.* 2009 Sep;30(9):1753-1762.)
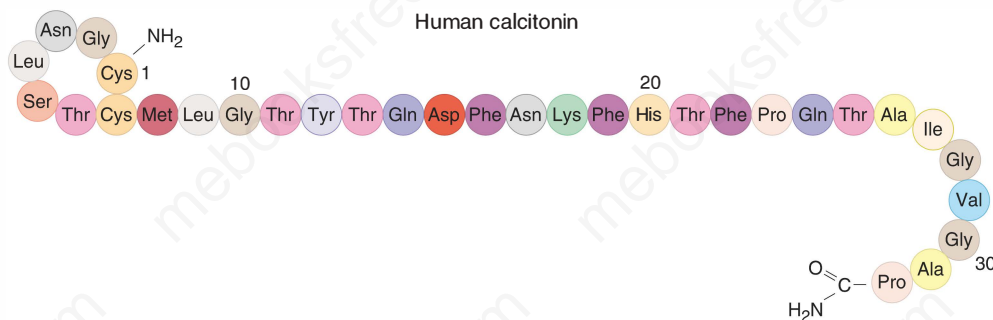
FIGURE 8–7 Amino acid sequence of human calcitonin, demonstrating its biochemical features, including an amino terminal disulfide bridge and carboxyl terminal prolinamide. (Reproduced with permission from McPhee SJ and Lingappa VR. *Pathophysiology of Disease: An Introduction to Clinical Medicine*. New York: McGraw-Hill; 1995.)

141-amino-acid precursor. In neurons of the central nervous system, CGRP is produced from a 128-amino-acid precursor. CGRP is a 37-amino-acid peptide with considerable homology to CT. The amino termini of both peptides incorporate a 7-member disulfide-bonded ring (see Figure 8–7). Acting through its own receptor, CGRP is among the most potent vasodilator substances known. When administered intravenously, CT produces a rapid and dramatic decline in levels of serum calcium and phosphate, primarily through actions on bone. The major effect of the hormone is to inhibit osteoclastic bone resorption. After exposure to CT, the morphology of the osteoclast changes rapidly. Within minutes, the cell withdraws its processes, shrinks in size, and retracts the ruffled border, the organelle of bone resorption, from the bone surface. Osteoclasts and cells of the proximal renal tubule express CT receptors. Like the PTH receptor, this is a receptor coupled through G_s alpha to adenylyl cyclase and thereby to the generation of cAMP in target cells. CT also has mild renal effects. In the kidney, CT inhibits the reabsorption of phosphate, thus promoting renal phosphate excretion. CT also induces a mild natriuresis and increases the renal excretion of calcium. The renal effects of CT are not essential for its acute effect on serum calcium levels, which results from blockade of bone resorption.

Although its secretory control by calcium and its antiresorptive actions enable CT to counter PTH in the control of calcium homeostasis, thus engendering bihormonal regulation, it is actually unlikely that CT plays an essential physiologic role in humans and other terrestrial animals. This is supported by two lines of evidence. First, removal of the thyroid gland—the major if not the only source of CT in mammals—has no perceptible impact on calcium handling or bone metabolism in postnatal life. Second, secretion of extremely high CT levels by medullary carcinoma of the thyroid (MCT), a malignancy of the C cell, likewise has no apparent effect on mineral homeostasis. Thus, in humans, CT is a hormone in search of an essential function. It plays a much more obvious homeostatic role in saltwater fish, in which the major challenge is maintenance of blood calcium levels in the sea, where the ambient calcium concentration of seawater is very high.

CT is of clinical interest for two reasons. First, serum CT is important as a tumor marker in MCT. Second, CT has

therapeutic uses as an inhibitor of osteoclastic bone resorption. CT can be administered either parenterally or as a nasal spray and is used in the treatment of Paget disease of bone, hypercalcemia, and osteoporosis.

VITAMIN D

Nomenclature

The term vitamin D (calciferol) refers to two secosteroids: vitamin D_2 (ergocalciferol) and vitamin D_3 (cholecalciferol) (Figure 8–8). Both are produced by photolysis from naturally occurring sterol precursors. Vitamin D_2 had been the principal form of vitamin D available for pharmaceutical purposes other than as dietary supplements, although this is changing as formulations with high doses of vitamin D_3 (eg, 10,000-50,000 units) are becoming more widely available (Table 8–2). Vitamin D_3 is produced from 7-dehydrocholesterol, a precursor of cholesterol found in high concentration in the skin. Although vitamin D_2 is metabolized somewhat differently than vitamin D_3, the active metabolites of vitamin D_2 and vitamin D_3 have equivalent biologic activity so the term "vitamin D," unless otherwise qualified, will be understood here to denote both. Vitamins D_2 and D_3 differ in their side chains: vitamin D_2 has a methyl group at C24 and a double bond at C22 to C23. These features alter the metabolism of vitamin D_2 compared with vitamin D_3; however, both are converted to 25-hydroxyvitamin D [25(OH)D] and 1,25-dihydroxyvitamin D [1,25(OH)$_2$D].

In the process of forming vitamins D_2 and D_3, the B ring of the sterol precursor is cleaved, and the A ring is rotated around the C5 to C6 double bond so that the 3β-hydroxyl group is positioned below the plane of the A ring. By convention, this hydroxyl group retains its designation as 3β, and the one position above the plane of the A ring is designated 1α. Hydroxylations in the side chain can lead to stereoisomers designated R and S. The natural position for the C24 hydroxyl group is R. Both the R and the S positions can be hydroxylated at C25 in the formation of 25,26(OH)$_2$D.

Because vitamin D can be formed *in vivo* (in the epidermis) in the presence of adequate amounts of ultraviolet light, it is more
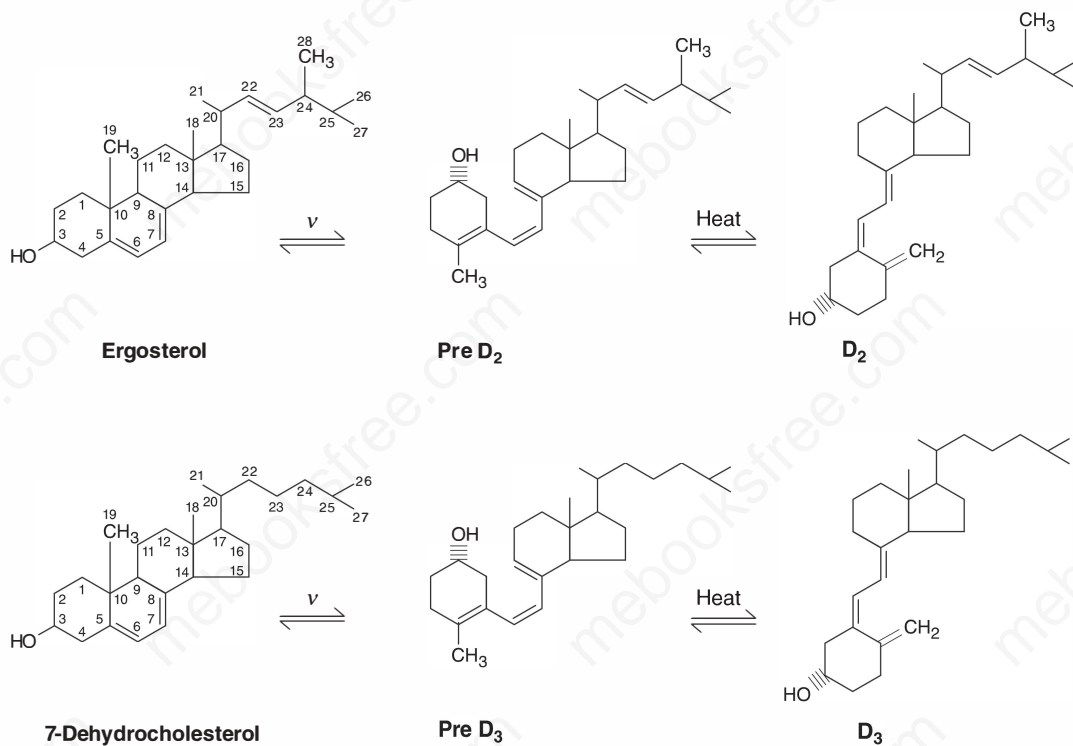
FIGURE 8–8 The photolysis of ergosterol and 7-dehydrocholesterol to vitamin D_2 (ergocalciferol) and vitamin D_3 (cholecalciferol), respectively. An intermediate is formed after photolysis (preD), which undergoes a thermal-activated isomerization to the final form of vitamin D. The rotation of the A-ring puts the 3-beta-hydroxyl group into a different orientation with respect to the plane of the A-ring during production of vitamin D.

properly considered a hormone (or prohormone) than a vitamin. To be biologically active, vitamin D must be metabolized further. The liver is the major but not the only organ that can metabolize vitamin D to its principal circulating form, 25(OH)D. The kidney and other tissues metabolize 25(OH)D to a variety of other metabolites, the most important of which are $1,25(OH)_2D$ and perhaps $24,25(OH)_2D$. A large number of other metabolites have been identified, but their physiologic roles are unclear. They may represent products destined only for elimination. Normal circulating levels of the principal metabolites are listed in Table 8–3. The recommended daily allowance for vitamin D in adults by the Institute of Medicine is 600 to 800 units (1 unit = 0.025 µg vitamin D), although the guidelines from the Endocrine Society recommend higher doses (eg, 2000 units) especially in older individuals who tend to have vitamin D deficiency or insufficiency more commonly.

TABLE 8–2 Commonly used vitamin D metabolites and analogs.

	Cholecalciferol, Ergocalciferol	Paricalcitol	Doxercalciferol	Calcitriol
Abbreviation	D_3, D_2	19-nor calcitriol	$1\alpha OHD_2$	$1,25(OH)_2D$
Physiologic dose	2.5-10 µg (1 ug = 40 units)			0.25-0.5 µg
Pharmacologic dose	0.625-5 mg	1-4 µg po; 2.5 µg/mL IV	0.5-2.5 µg po; 2 µg/mL IV	1-3 µg
Duration of action	1-3 mo	Hours to days	Hours to days	Hours to days
Clinical applications	Vitamin D deficiency Vitamin D malabsorption Hypoparathyroidism	Secondary hyperparathyroidism of chronic kidney disease	Secondary hyperparathyroidism of chronic kidney disease	Secondary hyperparathyroidism of chronic kidney disease Hypoparathyroidism Hypophosphatemic rickets with decreased calcitriol levels Acute hypocalcemia Pseudovitamin D deficient rickets Hereditary vitamin D resistant rickets

TABLE 8–3 Vitamin D and its metabolites.

Name	Abbreviation	Generic Name	Serum Concentration[a]
Vitamin D	D	Calciferol	1.6 ± 0.4 ng/mL
Vitamin D_3	D_3	Cholecalciferol	
Vitamin D_2	D_2	Ergocalciferol	
25-Hydroxyvitamin D	25(OH)D	Calcifediol	26.5 ± 5.3 ng/mL
1,25-Dihydroxyvitamin D	$1,25(OH)_2D$	Calcitriol	34.1 ± 0.4 pg/mL
24,25-Dihydroxyvitamin D	$24,25(OH)_2D$		1.3 ± 0.4 ng/mL
25,26-Dihydroxyvitamin D	$25,26(OH)_2D$		0.5 ± 0.1 ng/mL

[a]Values differ somewhat from laboratory to laboratory depending on the methodology used, sunlight exposure, and dietary intake of vitamin D in the population study. Children tend to have higher $1,25(OH)_2D$ levels than do adults.

Data from Bikle D. *Assay of Calcium Regulating Hormones*. New York: Springer; 1983.

Cutaneous Synthesis of Vitamin D

Vitamin D_3 is formed in the skin from 7-dehydrocholesterol, which is distributed throughout the epidermis and dermis but has its highest concentration in the lower layers of the epidermis, the stratum spinosum, and stratum basale. These epidermal layers also account for the highest production of vitamin D. The cleavage of the B ring of 7-dehydrocholesterol to form previtamin D_3 (see Figure 8–8) requires ultraviolet light with a spectrum of 280 to 320 nm (UVB). Following cleavage of the B ring, the previtamin D_3 undergoes thermal isomerization to vitamin D_3 but also to the biologically inactive compounds lumisterol and tachysterol. Formation of pre-D_3 is rapid and reaches a maximum within hours during exposure to solar or ultraviolet irradiation. The degree of epidermal pigmentation, the age of the skin, and the intensity of exposure all affect the time required to reach the maximum pre-D_3 concentration but do not alter that maximum. Continued ultraviolet light exposure then results in continued formation of the inactive compounds from pre-D_3. The formation of lumisterol is reversible, so lumisterol can be converted back to pre-D_3 as pre-D_3 levels fall. Short exposure to sunlight causes prolonged release of vitamin D_3 from the exposed skin because of the slow conversion of pre-D_3 to vitamin D_3 and the conversion of lumisterol to pre-D_3. Prolonged exposure to sunlight does not produce toxic quantities of vitamin D_3 because of the photoconversion of pre-D_3 to lumisterol and tachysterol.

Vitamin D_3 transport from the skin into the circulation has not been thoroughly studied. Vitamin D_3 is carried in the bloodstream primarily bound to vitamin D–binding protein (DBP), an α-globulin produced in the liver. DBP has a lower affinity for vitamin D_3 than other vitamin D metabolites such as 25(OH)D and $24,25(OH)_2D$. 7-Dehydrocholesterol, pre-D_3, lumisterol, and tachysterol bind to DBP even less well. Therefore, vitamin D_3 could be selectively removed from skin by the gradient established by selective binding to DBP. Because the deepest levels of the epidermis make the most vitamin D_3 when the skin is irradiated, the distance over which vitamin D_3 must diffuse to reach the circulation is short. However, simple diffusion is an unlikely means for so hydrophobic a molecule to enter the bloodstream. Epidermal lipoproteins may play a role in transport, but this remains to be established.

Dietary Sources and Intestinal Absorption

Dietary sources of vitamin D are clinically important because exposure to ultraviolet light may not be sufficient to maintain adequate production of vitamin D in the skin. The farther away from the equator one lives, the shorter the period of the year during which the intensity of sunlight is sufficient to produce vitamin D_3. Most milk products in the United States are supplemented with vitamin D. Unfortified milk products contain little or no vitamin D. Although plants and mushrooms contain ergosterol, their content of vitamin D_2 is limited unless they are irradiated with ultraviolet light during processing. Vitamin D is found in moderate to high concentrations in fish oils and fish liver and in lesser concentrations in eggs. Vitamin D is absorbed from the diet in the small intestine with the help of bile salts. Drugs that bind bile salts such as colestipol and various malabsorption syndromes reduce vitamin D absorption. Most of the vitamin D passes into the lymph in chylomicrons, but a significant amount is absorbed directly into the portal system. The presence of fat in the lumen decreases vitamin D absorption. 25(OH)D and calcitriol are preferentially absorbed into the portal system and are less influenced by the amount of fat in the lumen. Biliary conjugates of the vitamin D metabolites have been identified, and an enterohepatic circulation of these metabolites has been established. Vitamin D is taken up rapidly by the liver and is metabolized to 25(OH)D. 25(OH)D is also transported in the blood bound to DBP. Little vitamin D is stored in the liver. Excess vitamin D is stored in adipose tissue and muscle.

Binding Proteins for Vitamin D Metabolites

As noted earlier, vitamin D metabolites are transported in the blood bound principally to DBP (85%) and albumin (15%). DBP binds 25(OH)D and $24,25(OH)_2D$ with approximately 30 times greater affinity than it binds $1,25(OH)_2D$ and vitamin D. In most species including humans, DBP binds the D_3 metabolites with greater affinity than the D_2 metabolites, which alter their pharmacokinetics. DBP circulates at a concentration (5×10^{-6} mol/L) approximately 50 times greater than the total concentrations of the vitamin D metabolites. DBP levels are reduced in

liver disease and in the nephrotic syndrome and increased during pregnancy and with estrogen administration, but are not altered by states of vitamin D deficiency or excess. The high affinity of DBP for the vitamin D metabolites and the large excess-binding capacity maintain the free and presumed biologically active concentrations of the vitamin D metabolites at very low levels—approximately 0.03% and 0.4% of the total 25(OH)D and 1,25(OH)$_2$D levels, respectively. Liver disease reduces the total level of the vitamin D metabolites commensurate with the reduction in DBP and albumin levels, but the free concentrations of the vitamin D metabolites remain normal in most subjects. This fact must be borne in mind when evaluating a patient with liver disease to determine whether such a patient is truly vitamin D deficient. Pregnancy, on the other hand, increases both the free and total concentrations by increasing DBP levels, altering the binding of the metabolites to DBP, and increasing the production of 1,25(OH)$_2$D. The binding of the vitamin D metabolites to DBP appears to occur at the same site, and saturation of this site by one metabolite can displace the other metabolites. This is an important consideration in vitamin D toxicity because the very high levels of 25(OH)D that characterize this condition displace the often normal levels of 1,25(OH)$_2$D, leading to elevated free and biologically active concentrations of 1,25(OH)$_2$D. This phenomenon at least partially explains the hypercalcemia and hypercalciuria that mark vitamin D intoxication even in patients with normal total 1,25(OH)$_2$D levels.

At present, it is not clear whether DBP functions just to maintain a circulating reservoir of vitamin D metabolites or whether DBP participates in transporting the vitamin D metabolites to their target tissues. However, some tissues, such as the kidney and parathyroid gland, express megalin and cubilin that enable DBP-bound metabolites to be transported into cells. A different set of proteins, related to heat shock proteins, are found within cells that bind the vitamin D metabolites and appear to facilitate their further metabolism. These proteins are distinct from DBP and the nuclear hormone receptors for calcitriol, which will be discussed subsequently.

Metabolism

The conversion of vitamin D to 25(OH)D occurs principally in the liver (Figure 8–9). Both mitochondria and microsomes have the capacity to produce 25(OH)D but with different kinetics and with different enzymes. The mitochondrial enzyme has been identified as the 5-β–cholestane-3α,7α,12α-triol-27(26)-hydroxylase (CYP27A1), a cytochrome P450 mixed-function oxidase, and its cDNA has been isolated. A number of microsomal P450 enzymes have been shown to have 25-hydroxylase activity with variable preference for D$_2$ or D$_3$ and their 1-hydroxylated metabolites. The microsomal 25-hydroxylase that has received the most attention is CYP2R1. Like the mitochondrial enzyme, these are not specific for vitamin D. Regulation of 25(OH)D
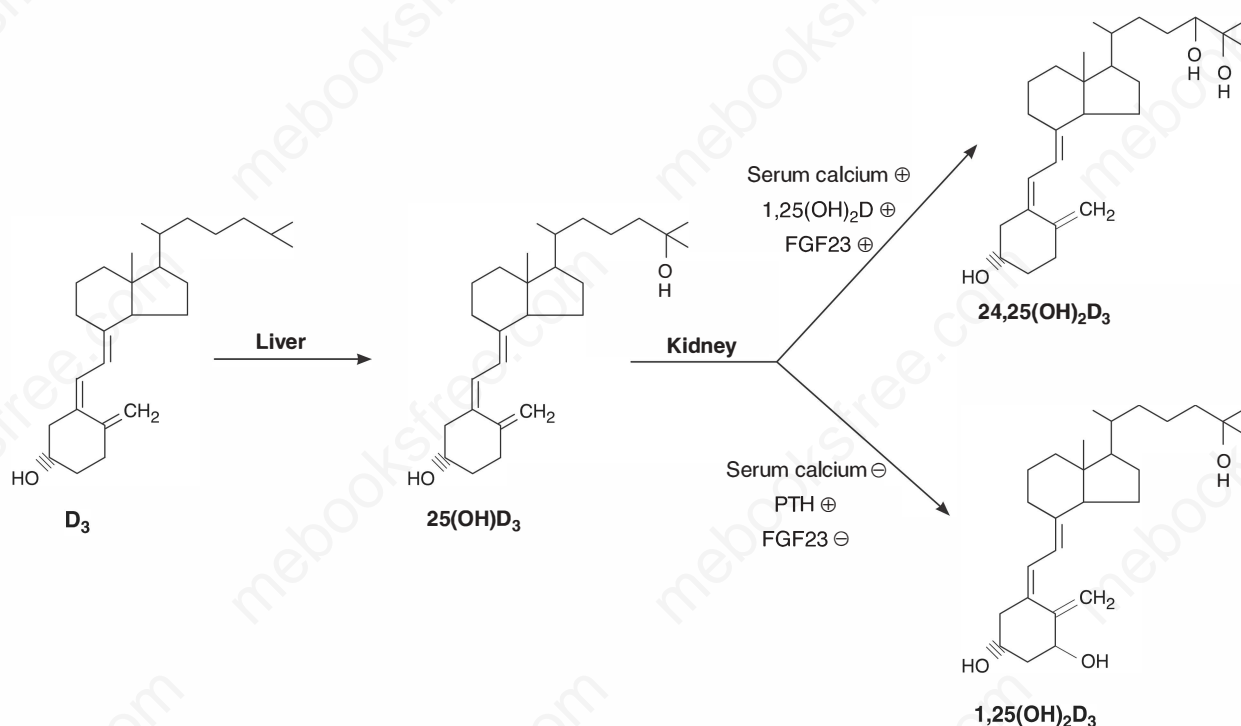
FIGURE 8–9 The metabolism of vitamin D. The liver converts vitamin D to 25(OH)D. The kidney converts 25(OH)D to 1,25(OH)$_2$D and 24,25(OH)$_2$D. Control of metabolism is exerted primarily at the level of the kidney where low serum phosphate, low serum calcium, and high parathyroid hormone (PTH) levels favor production of 1,25(OH)$_2$D whereas FGF23, high serum calcium and phosphate, and 1,25(OH)$_2$D inhibit 1,25(OH)$_2$D production while increasing 24,25(OH)$_2$D production. Plus (+) and minus (−) signs denote the stimulatory and inhibitory enzymatic reactions, respectively, driving the metabolic steps indicated.

production is difficult to demonstrate. Drugs such as phenytoin and phenobarbital reduce serum 25(OH)D levels, primarily by increased catabolism of 25(OH)D and vitamin D. Liver disease leads to reduced total (but not necessarily free) serum 25(OH)D levels, primarily because of reduced DBP and albumin synthesis and not reduced synthesis of 25(OH)D. Vitamin D deficiency is marked by low blood levels of 25(OH)D, primarily because of lack of substrate for the 25-hydroxylase. Vitamin D intoxication, on the other hand, leads to increased levels of 25(OH)D because of the lack of feedback inhibition on the 25-hydroxylases.

Control of vitamin D metabolism is exerted principally in the kidney (see Figure 8–9). $1,25(OH)_2D$ and $24,25(OH)_2D$ are produced by cytochrome mixed-function oxidases in mitochondria of the proximal tubules. cDNAs encoding these enzymes have both been identified and show substantial homology to each other and to the mitochondrial 25-hydroxylase (CYP27A1). Their homology to other mitochondrial steroid hydroxylases

is considerable but not as high. Both 24-hydroxylase and 1-hydroxylase are widely distributed. Their roles in extra-renal tissues appear to be paracrine or autocrine, with limited contribution to circulating levels of their metabolic products.

The kidney remains the principal source of circulating $1,25(OH)_2D$. Production of $1,25(OH)_2D$ in the kidney is stimulated by PTH and insulin-like growth factor-1 (IGF-1) and is inhibited by fibroblast-derived growth factor 23 (FGF23) and high blood levels of calcium and phosphate. Production of FGF23, a recently identified hormone, occurs primarily in bone and is regulated by the levels of serum phosphate and $1,25(OH)_2D$ (Figure 8–10). FGF23 inhibits $1,25(OH)_2D$ production, blocks renal phosphate reabsorption, and acts to lower serum phosphate (see Figure 8–10 and Chapter 21).

Calcium and phosphate also have both direct and indirect effects on $1,25(OH)_2D$ production. In addition to direct actions on the 1-hydroxylase to reduce its activity, calcium inhibits PTH
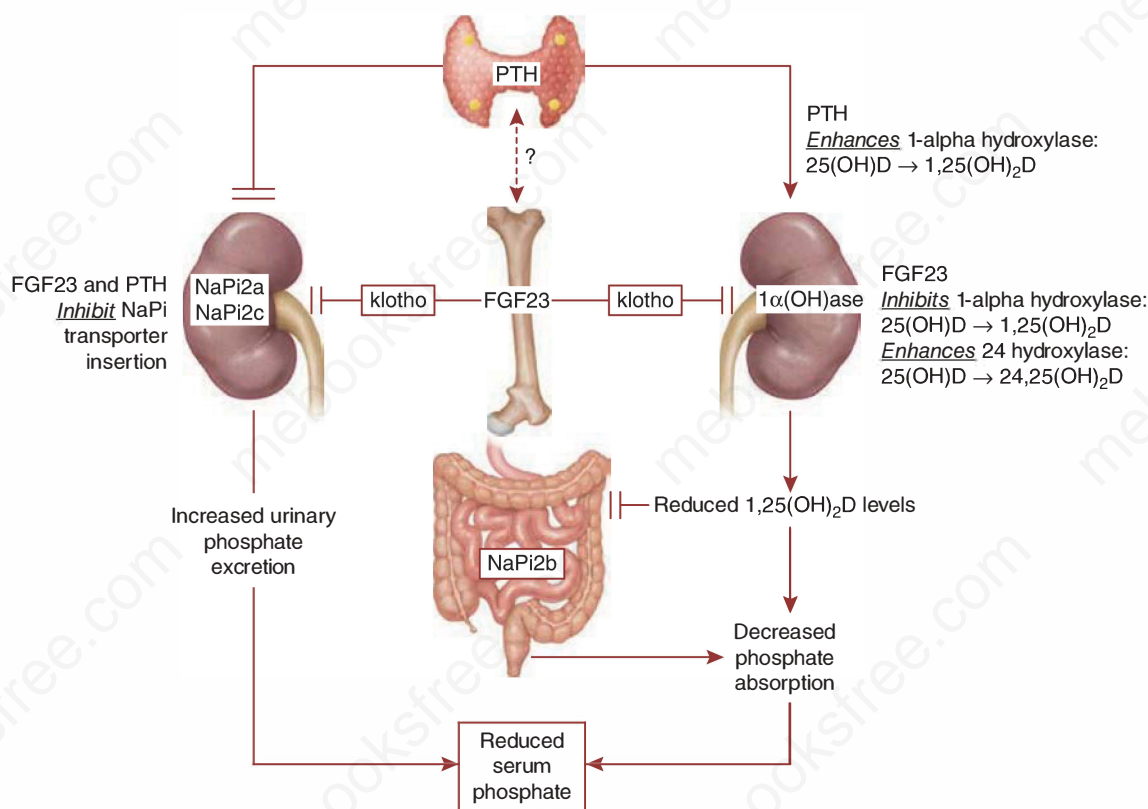
FIGURE 8–10 A simplified schema for the regulation of serum phosphate by PTH, $1,25(OH)_2D$ and fibroblast growth factor 23 (FGF23). FGF23 is produced predominantly by osteocytes buried deep in the bone matrix. Its actions are mediated by FGF receptors, differing in their expression in different tissues, and by the membrane co-receptor molecule alpha klotho. The alpha klotho/FGF receptor complex mediates high-affinity activation of signal transduction by FGF23 in key target tissues. FGF23 has potent effects on calcium and phosphate homeostasis as shown. FGF23 reduces $1,25(OH)_2D$ production by reducing 1-alpha hydroxylase activity and expression and increases its catabolism by enhancing 24-hydroxylase activity and expression. Together this reduces levels of $1,25(OH)_2D$, which has downstream effects to reduce intestinal calcium and phosphate absorption. FGF23 also reduces the expression of key phosphate transporters in the kidney (NaPi2a and NaPi2c) thereby facilitating the excretion of phosphate. The reciprocal interaction between PTH and FGF23 is unclear and may depend on the physiologic circumstances, but PTH independently also stimulates phosphate excretion via the same NaPi transporters. The combined actions of FGF23 lower serum phosphate concentrations. Not displayed here are the negative feedback control mechanisms including: the ability of $1,25(OH)_2D$ and phosphate to stimulate FGF23 production in bone and the suppression of PTH secretion by $1,25(OH)_2D$. Excessive FGF23 production and/or action (hypophosphatemia, chronic phosphate wasting) can result in abnormal bone matrix mineralization, causing rickets or osteomalacia.

secretion by the parathyroid glands, and this lowers renal $1,25(OH)_2D$ production. Phosphate inhibits the production of growth hormone by the pituitary gland, leading to decreased IGF-1 generation, and stimulates FGF23 production, each step contributing to the indirect inhibition of $1,25(OH)_2D$ production by phosphate.

$1,25(OH)_2D$ inhibits its own synthesis at least in the kidney (but not in other tissues) and stimulates the production of $24,25(OH)_2D$ and $1,24,25(OH)_3D$ by the 24-hydroxylase, which are the initial products of both $25(OH)D$ and $1,25(OH)_2D$ catabolism, respectively. Other hormones such as prolactin and CT may regulate $1,25(OH)_2D$ production, but their roles under normal physiologic conditions in humans have not been demonstrated.

Extrarenal production of $1,25(OH)_2D$ is regulated in a cell-specific fashion. Cytokines such as γ-interferon (γ-IFN) and tumor necrosis factor-α (TNF-α) stimulate $1,25(OH)_2D$ production by macrophages and keratinocytes. $1,25(OH)_2D$ and calcium have minimal effects on $1,25(OH)_2D$ production by these cells. This is important in understanding the pathophysiology of hypercalcemia and the increased $1,25(OH)_2D$ levels in patients with sarcoidosis, lymphomas, and other granulomatous diseases, in that feedback regulation by calcium and $1,25(OH)_2D$ does not occur.

Mechanisms of Action

The main function of vitamin D metabolites is the regulation of calcium and phosphate homeostasis, which occurs in conjunction with PTH. The gut, kidney, and bone are the principal target tissues for this regulation. The major pathologic complication of vitamin D deficiency is rickets (in children with open epiphyses) or osteomalacia (in adults), which results mainly from the deficiency of calcium and phosphate required for bone mineralization. $1,25(OH)_2D$ is the most biologically active, if not the only vitamin D metabolite involved in maintaining calcium and phosphate homeostasis.

Most of the cellular processes regulated by $1,25(OH)_2D$ involve the nuclear vitamin D receptor (VDR), a 50-kDa protein related by structural and functional homology to a large class of nuclear hormone receptors including the steroid hormone receptors, thyroid hormone receptors (TRs), retinoic acid and retinoid X receptors (RARs, RXRs), and endogenous metabolite receptors, including peroxisome proliferation activator receptors (PPARs), liver X receptors (LXRs), and farnesoid X receptors (FXRs) (see Chapter 1). These receptors are transcription factors. VDRs, as is the case with TRs, RARs, PPARs, LXRs, and FXRs, generally act by forming heterodimers with RXRs. The VDR–RXR complex then binds to specific regions within the regulatory portions of the genes whose expression is controlled by $1,25(OH)_2D$ (Figure 8–11). The regulatory regions are called vitamin D response elements (VDREs) and are generally composed of two stretches of six nucleotides (each called a half-site) with a specific and nearly identical sequence separated by three nucleotides with a less specific sequence (a direct repeat with three base-pair separation; DR3). Numerous exceptions to this rule can be found such as a
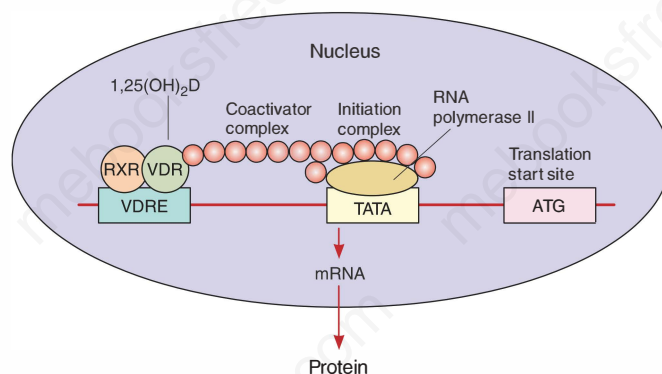
FIGURE 8–11 $1,25(OH)_2D$-initiated gene transcription. $1,25(OH)_2D$ enters the target cell and binds to its receptor, VDR. The VDR then heterodimerizes with the retinoid X receptor, RXR. This increases the affinity of the VDR–RXR complex for the VDRE, a specific sequence of nucleotides in the promoter region of a vitamin D responsive gene. Binding of the VDR–RXR complex to the VDRE attracts a complex of proteins termed coactivators to the VDR–RXR complex, some of which open up the gene for transcription by acetylating the histones and other proteins which communicate with and activate the RNA polymerase II complex. Transcription of the gene is initiated to produce the corresponding mRNA, which leaves the nucleus for the cytoplasm to be translated to the corresponding protein.

VDR–RAR complex binding to a DR6 (direct repeat with six base-pair separation) or inverted palindromes (repeat elements going in opposite directions) with various degrees of spacing (eg, IP9). Other nuclear hormone receptors bind to similar elements but with different spacing or orientation of the half-sites. The binding of the VDR–RXR complex to the VDRE then attracts a number of other proteins called coactivators, which may function to open up the gene for transcription by acetylating the histones covering the gene or by bridging the gap between the VDRE and the initiation complex containing the RNA polymerase to signal the beginning of transcription. Not all actions of $1,25(OH)_2D$, however, can be explained by changes in gene expression. VDRs in the membrane and a membrane-bound protein identified as protein disulfide isomerase A3 (Pdia3), initially known as the $1,25D_3$-membrane–associated rapid response steroid binding protein ($1,25D_3$-MARRS), may mediate the rapid effects of $1,25(OH)_2D$ on calcium influx and protein kinase C activity observed in a number of tissues, effects too rapid to be explained by genomic activity.

A. Intestinal calcium and phosphate transport   Intestinal calcium transport is the best understood target tissue response of $1,25(OH)_2D$. Calcium transport through the intestinal epithelium proceeds by at least three distinct steps: (1) entrance into the cell from the lumen across the brush border membrane down a steep electrochemical gradient; (2) passage through the cytosol, probably within subcellular organelles such as mitochondria and endosomes; and (3) removal from the cell against a steep electrochemical gradient at the basolateral membrane.

Each of these steps is regulated by 1,25(OH)$_2$D. At the brush border, 1,25(OH)$_2$D induces a change in the binding of calmodulin to brush border myosin 1, a unique form of myosin found only in the intestine, where it resides primarily in the microvillus bound to actin and to the plasma membrane. The calmodulin-myosin 1 complex may provide the mechanism for removing calcium from the brush border after it crosses the membrane into the cell. Changes in the phospholipid composition of the brush border may explain the increased flux of calcium across this membrane after calcitriol administration. None of these changes requires new protein synthesis. On the other hand, a Ca^{2+} channel, the transient receptor potential cation channel 6 (TRPV6) in the brush border membrane, is induced by 1,25(OH)$_2$D and is likely the major mechanism by which calcium enters the intestinal epithelial cell. The transport of calcium through the cytosol requires a vitamin D-inducible protein called calbindin. Calbindin exists either in a 28-kDa or 9-kDa form, depending on the species and tissue. Calbindin has a high affinity for calcium. If its synthesis is blocked, the calcium content of the cytosol and mitochondria increases, and efficiency of transport is reduced. At the basolateral membrane, calcium is removed from the cell by an ATP-driven pump, the Ca^{2+}-ATPase (PMCA1b), a protein also induced by 1,25(OH)$_2$D. Recent studies, however, in which calbindin or TRPV6 expression is deleted show surprisingly little impact on intestinal calcium transport (calbindin knockout) or only a partial reduction (TRPV6 knockout) suggesting that there is more to learn about the mechanisms governing intestinal calcium transport. Nevertheless in mice in which the VDR is rendered nonfunctional (VDR knockout), intestinal calcium transport is markedly impaired, indicating the importance of VDR function—presumably via regulation of gene expression—in governing intestinal calcium transport.

The regulation of phosphate absorption from the intestine is less well understood. It occurs primarily in the jejunum. The sodium phosphate transporter, NaPiIIb, a member of the family of phosphate transporters that includes NaPiIIa and IIc in the kidney, is thought to be the main mechanism by which phosphate is absorbed from the luminal contents of the intestine. 1,25(OH)$_2$D increases the levels of NaPiIIb in the apical (luminal) membrane, but a direct stimulation of NaPiIIb expression has not been clearly shown. Moreover, in states of dietary phosphate restriction, phosphate absorption is increased even in animals lacking either the VDR or Cyp27b1, indicating that vitamin D independent factors also control phosphate absorption.

B. Actions of vitamin D in bone   The critical role of 1,25(OH)$_2$D in regulating bone formation and resorption is evidenced by the development of rickets in children who are vitamin D-deficient, have inactivating mutations in *CYP27B1* (the 1-hydroxylase) (pseudovitamin D-deficiency rickets), or lack a functioning VDR [hereditary 1,25(OH)$_2$D-resistant rickets]. The histologic and radiologic appearance of bone in these genetic conditions is quite similar to that in patients with severe vitamin D deficiency.

Patients with inactivating mutations in the *1-hydroxylase* can be treated successfully with calcitriol (but not with vitamin D), whereas those with mutations in *VDR* are resistant to calcitriol as well as to vitamin D. These experiments of nature, however, do not exclude the possibility that other vitamin D metabolites, in combination with 1,25(OH)$_2$D, are essential for normal bone metabolism.

In vitamin D deficiency, all vitamin D metabolites are reduced, but 25(OH)D and 24,25(OH)$_2$D tend to be reduced out of proportion to 1,25(OH)$_2$D, which may be maintained in the low normal range despite clear abnormalities in bone mineralization. Thus, it is reasonable to ask whether 1,25(OH)$_2$D is the only vitamin D metabolite of consequence for the regulation of skeletal homeostasis. In fact, a number of studies point to unique actions of 24,25(OH)$_2$D that cannot be replicated by 1,25(OH)$_2$D, especially in cartilage, although this remains controversial. Studies in mice, in which the *24-hydroxylase* has been inactivated (knocked-out), support the concept that 24,25(OH)$_2$D has a role in bone formation that is not replaced by 1,25(OH)$_2$D. Despite those observations, many of the metabolic and skeletal abnormalities in 24-hydroxylase–null animals (including humans) may be explained by their inability to regulate 1,25(OH)$_2$D levels. Nevertheless, the need for more than just 1,25(OH)$_2$D is an important concept in the treatment of vitamin D deficiency; this is best done with vitamin D or 25(OH)D rather than calcitriol alone, because vitamin D serves as a precursor for both 1,25(OH)$_2$D and 24,25(OH)$_2$D.

The mechanisms by which 1,25(OH)$_2$D regulates skeletal homeostasis remain uncertain. Provision of adequate calcium and phosphate for mineralization is clearly important. The rickets present in patients with defective VDRs can be cured with calcium infusions. Similarly, in mice in which the VDR is rendered nonfunctional (knocked-out), rickets is prevented with diets high in calcium and phosphate. Thus, even though the osteoblast, the cell responsible for forming new bone, expresses VDR, and the transcription of a number of proteins in bone is regulated by 1,25(OH)$_2$D, it is not clear how essential the direct actions of 1,25(OH)$_2$D on bone are for bone formation and mineralization. Thus, bone is a good example of a tissue in which 1,25(OH)$_2$D and calcium work together, such that limited quantities of one may be compensated for by increased amounts of the other. However, recent studies suggest that although 1,25(OH)$_2$D promotes bone formation in situations of adequate calcium, it inhibits bone formation when the 1,25(OH)$_2$D levels are high, but calcium levels are low.

In organ cultures of bone, the best established action of 1,25(OH)$_2$D is to stimulate bone resorption. This is accompanied by an increase in osteoclast number and activity and decreased collagen synthesis. The increase in osteoclast number and activity is now known to be mediated by the production in osteoblasts of a membrane-bound protein called receptor activator of NF-kappa B ligand (RANKL), which acts on its receptor (RANK) in osteoclasts and their precursors to stimulate osteoclast differentiation and activity (discussed later). 1,25(OH)$_2$D is one of several hormones (PTH and selected cytokines are others) that stimulate RANKL production. Vitamin D deficiency does not decrease

bone resorption, probably because PTH levels are elevated. High levels of PTH could, therefore, compensate for the lack of $1,25(OH)_2D$ at the level of RANKL production. The presence of partial skeletal resistance to PTH in patients with vitamin D deficiency may be explained by this mechanism.

$1,25(OH)_2D$ also promotes the differentiation of osteoblasts. This action is less well-delineated than the promotion of osteoclast differentiation. Osteoblasts pass through a well-defined sequence of biochemical processes as they differentiate from proliferating osteoprogenitor cells to cells capable of producing and mineralizing matrix. The effect of $1,25(OH)_2D$ on the osteoblast depends on the stage of differentiation at which the $1,25(OH)_2D$ is administered. Early in the differentiation process, in cells characterized as immature osteoblast-like cells, $1,25(OH)_2D$ stimulates collagen production and alkaline phosphatase activity. These functions are inhibited in more mature osteoblasts. On the other hand, osteocalcin production is stimulated by $1,25(OH)_2D$ only in mature osteoblasts. Experiments *in vivo* demonstrating that excessive $1,25(OH)_2D$ can inhibit normal bone mineralization—leading to the paradoxical appearance of osteomalacia—can be understood in the context of these differential effects of $1,25(OH)_2D$ on the osteoblast as it differentiates. Excess $1,25(OH)_2D$ can disrupt the differentiation pathway.

C. Actions of vitamin D in kidney   The kidney expresses VDRs, and $1,25(OH)_2D$ stimulates the expression of calbindin, TRPV5 (the renal homologue of the intestinal TRPV6), and Ca^{2+}-ATPase (PMCA) in the distal tubule as well as $24,25(OH)_2D$ production in the proximal tubule. However, the role of $1,25(OH)_2D$ in regulating calcium and phosphate transport across the renal epithelium remains controversial. $25(OH)D$ may be more important than $1,25(OH)_2D$ in acutely stimulating calcium and phosphate reabsorption by the kidney tubules. *In vivo* studies are complicated by the effect of $1,25(OH)_2D$ on other hormones, particularly PTH, which appears to be more important than the vitamin D metabolites in regulating calcium and phosphate handling by the kidney.

D. Actions of vitamin D in other tissues   An exciting observation is that VDRs are found in a large number of tissues beyond the classic target tissues—gut, bone, and kidney—and these tissues respond to $1,25(OH)_2D$. Furthermore, many of these tissues express *CYP27B1* (the 1-hydroxylase) and so are capable of producing their own $1,25(OH)_2D$ from circulating $25(OH)D$. These tissues include elements of the hematopoietic and immune systems; cardiac, skeletal, and smooth muscle; and brain, liver, breast, endothelium, skin (keratinocytes, melanocytes, and fibroblasts), and endocrine glands (pituitary, parathyroid, pancreatic islets [beta cells], adrenal cortex and medulla, thyroid, ovary, and testis). Furthermore, malignancies developing in these tissues often contain VDRs and respond to the antiproliferative actions of $1,25(OH)_2D$.

The responses of these tissues to $1,25(OH)_2D$ are as varied as the tissues themselves. $1,25(OH)_2D$ regulates hormone production and secretion, including stimulation of insulin secretion from the pancreas and prolactin from the pituitary, but inhibiting PTH secretion from the parathyroid gland, renin secretion from the kidney, and atrial naturetic peptides from the heart. $1,25(OH)_2D$ enables the innate immune system by inducing antimicrobial peptides such as cathelicidin but suppresses the adaptive immune response by blocking the maturation of dendritic cells and shifting the balance of T-helper cell differentiation from Th1 and Th17 to Th2 and Treg. Myocardial contractility and vascular tone are modulated by $1,25(OH)_2D$. $1,25(OH)_2D$ reduces the rate of proliferation of many cell lines, including normal keratinocytes, fibroblasts, lymphocytes, and thymocytes as well as abnormal cells of mammary, skeletal, intestinal, lymphatic, and myeloid origin. Differentiation of numerous normal cell types, including keratinocytes, lymphocytes, hematopoietic cells, intestinal epithelial cells, osteoblasts, and osteoclasts as well as abnormal cells of the same lineage is enhanced by $1,25(OH)_2D$. Thus, the potential for manipulating a vast array of physiologic and pathologic processes with calcitriol and its analogs is enormous. This promise is being realized in that vitamin D analogs are being used to treat psoriasis, uremic hyperparathyroidism, and osteoporosis. Epidemiologic and animal studies indicate a role for vitamin D in the prevention and treatment of both type 1 and type 2 diabetes mellitus, autoimmune diseases such as multiple sclerosis and inflammatory bowel disease, and infections such as tuberculosis. Trials of vitamin D analogs in the treatment of a variety of cancers are also being conducted. Thus, although regulation of bone mineral homeostasis remains the major physiologic function of vitamin D and its metabolites, clinical applications for these compounds and their analogs are being evaluated outside the classic target tissues.

HOW VITAMIN D AND PTH CONTROL MINERAL HOMEOSTASIS

Consider a person who switches from a high normal to a low normal intake of calcium and phosphate—from 1200 mg/d to 300 mg/d of calcium (the equivalent of leaving three glasses of milk out of the daily diet). The net absorption of calcium falls sharply, causing a transient decrease in the serum calcium level. The homeostatic response to this transient hypocalcemia is led by an increase in PTH, which stimulates the release of calcium and phosphate from bone and the retention of calcium by the kidney. The phosphaturic effect of PTH allows elimination of phosphate, which is resorbed from bone together with calcium. In addition, the increase in PTH, along with the fall in serum calcium and serum phosphate, activates renal $1,25(OH)_2D$ synthesis. In turn, $1,25(OH)_2D$ increases the fractional absorption of calcium and further increases bone resorption. External calcium balance is thus restored by increased fractional absorption of calcium and increased bone resorption at the expense of higher steady state levels of PTH and $1,25(OH)_2D$.

MEDULLARY CARCINOMA OF THE THYROID

MCT, a neoplasm of thyroidal C cells, accounts for 3% to 5% of all thyroid malignancies. Approximately 75% of MCTs are sporadic. The remainder is familial and associated with one of three heritable syndromes: familial isolated MCT; multiple endocrine neoplasia 2A (MEN 2A), consisting of MCT, pheochromocytoma, and primary hyperparathyroidism; or multiple endocrine neoplasia 2B (MEN 2B), consisting of MCT, pheochromocytoma, multiple mucosal neuromas and, very rarely, primary hyperparathyroidism (Table 8–4). The MEN syndromes are more extensively discussed in Chapter 22.

The pathogenesis of hereditary MCT is largely due to the consequences of mutations in the *RET* proto-oncogene located on chromosome 10q11.2. *RET* (rearranged during transfection) encodes a membrane tyrosine kinase receptor whose ligands belong to the glial cell line-derived neurotrophic factor (GDNF) family. This receptor is expressed developmentally in migrating neural crest cells that give rise to hormone-secreting neuroendocrine cells (eg, C cells and adrenal medullary cells) and to the parasympathetic and sympathetic ganglia of the peripheral nervous system. Remarkably, different mutations in *RET* can produce five distinct diseases (see Table 8–4). Inheritance of certain activating mutations in *RET* is responsible for MEN 2A and familial MCT. Inheritance of a different set of activating mutations causes MEN 2B. In over half of sporadic MCTs, the tumor has a clonal somatic mutation (present in the tumor but not in genomic DNA), which is identical to one of the mutations responsible for the familial forms of MCT. These somatic mutations clearly play a causative role in sporadic MCT. Mutations in *RET* can also produce Hirschsprung disease, a congenital absence of the enteric parasympathetic ganglia, that can lead to a disturbance in intestinal motility, resulting in megacolon.

TABLE 8–4 Clinical presentations resulting from RET mutations.

Disorder	Characteristic Components
MEN 2A	MCT, pheochromocytoma, primary hyperparathyroidism
MEN 2A with lichen amyloidosis	MEN 2A + pruritic lesions of upper back
Familial MCT	MCT
MEN 2A or Familial MCT and Hirschsprung disease	MCT, Hirschsprung disease
MEN 2B	MCT, pheochromocytoma, marfanoid habitus, intestinal and mucosal ganglioneuromatosis

Modified with permission from Brandi ML, et al. Guidelines for diagnosis and therapy of MEN type 1 and type 2. *J Clin Endocrinol Metab*. 2001 Dec;86(12): 5658-5671.

MCT is usually located in the middle or upper portions of the thyroid lobes. It is typically unilateral in sporadic cases but often multicentric and bilateral in familial forms of the disease. The natural history of MCT is variable. Sporadic tumors may be quite aggressive or very indolent; the mean 5-year survival rate is about 50%. The behavior of familial forms varies among syndromes. MEN 2B has the most aggressive form of MCT, with a 2-year survival of about 50%; MEN 2A has a course similar to that of sporadic forms of this tumor, and familial MCT has the most indolent course of all MCT. The natural history of the tumor in familial MCT, MEN 2A, and MEN 2B syndromes has been dramatically impacted by early prophylactic total thyroidectomy that is now advised in most carriers of disease-causing *RET* mutations. MCT may spread to regional lymph nodes or undergo hematogenous spread to the lungs and other viscera. When metastatic, this tumor is sometimes associated with a chronic diarrhea syndrome. The pathogenesis of the diarrhea is unclear. In addition to CT, these tumors secrete a variety of other bioactive products, including prostaglandins, serotonin, histamine, and peptide hormones (ACTH, somatostatin, corticotropin-releasing hormone). In some cases, the associated diarrhea responds dramatically to treatment with a long-acting somatostatin analog such as octreotide, which blocks secretion of these bioactive products.

CT is a tumor marker for MCT. It is most sensitive for this purpose when secretion is stimulated with provocative agents. Provocative tests have used pentagastrin (0.5 ug/kg intravenously over 5 seconds) or a rapid infusion of calcium gluconate (2 mg calcium/kg over 1 minute). Blood samples are obtained at baseline and 1, 2, and 5 minutes after the stimulus. For maximal sensitivity, both tests are usually combined, with the calcium infusion immediately followed by administration of pentagastrin. Although basal CT levels are often normal in early tumors, CT levels may be many times higher than normal in patients with disseminated MCT. Despite this, the patients are uniformly normocalcemic. Although tumors secrete larger molecular weight forms of CT with decreased biologic activity, monomeric CT levels are often high as well. *RET* oncogene analysis has replaced provocative testing in familial MCT, although basal CT levels are still used to follow disease burden postoperatively and during chemotherapy trials.

Members of families carrying *RET* mutations must be screened for MCT and, if positive, for the associated tumors that occur in MEN 2A and MEN 2B (see Table 8–4 and Chapter 22). In the case of MCT, the presence of the *RET* mutation in an individual generally leads to the recommendation for a total thyroidectomy prior to the development of frank malignancy or abnormal CT levels. It is recommended in a kindred with a known *RET* mutation that children be screened at birth for the carrier state. Total thyroidectomy with or without central compartment lymph node dissection can be performed in mutation carriers, and further testing can be discontinued in genetically normal family members. The timing of prophylactic thyroidectomy in carriers of *RET* mutations, however, depends on the genotype present (ie, codon mutated in *RET*) (Table 8–5). As supported by the genotype-phenotype correlations in these diseases, surgery is recommended

TABLE 8–5 American Thyroid Association risk categories for various RET mutations.

Mutation			MEN 2A			MEN 2B		
Exon	Codon	FMCT	MCT	HPT	PHEO	MCT	PHEO	ATA Risk Level
5	321	+	***					A
8	515	+	***					A
	533	+	***		*			A
10	609	+	***	**	*			B
	611	+	***	**	*			B
	618	+	***	**	**			B
	620	+	***	**	**			B
11	630	+	***	*	*			B
	631	+						B
	634	+	***	**	***			C
	634 bp dup	---	***	**				B
	649	+	**	*				A
	666	+	***/**		**			A
13	768	+	***	*	*			A
	777	+	**					A
	790	+	***	*	**/*			A
	791	+	***	**	**			A
14	804	+	***	**/*	*			A
	804+805	---				***	***	D
	804+806	---				***	***	D
	804+904	---		**		***		D
	819	+						A
	833	+						A
	844	+						A
15	866	+	***					A
	883	---				***	***	D
	891	+	***	*	*			A
16	912	+	**					A
	918	---				***	***	D

Abbreviations: ATA, American Thyroid Association; MEN, multiple endocrine neoplasia; FMCT, familial medullary carcinoma of thyroid; MCT, medullary carcinoma of thyroid; HPT, primary hyperparathyroidism; PHEO, pheochromocytoma; dup, duplication.

Penetrance for the disorder:

***most cases, **few cases, *rare.

A, may delay surgery beyond 5 years of age if certain criteria are met; B, consider surgery before age 5 years and may delay if certain criteria are met; C, surgery before 5 years of age; D, surgery as soon as possible within the first year of life. Consultation with current ATA Guidelines for the most updated information is recommended. Criteria used to consider surgical timing are having a normal annual basal and/or stimulated serum calcitonin level and cervical ultrasound and possessing a family history of less aggressive MCT.

Data from American Thyroid Association Guidelines Task Force, Kloos RT, Eng C, et al. Medullary thyroid cancer: management guidelines of the American Thyroid Association. *Thyroid.* 2009 Jun;19(6):565-612.

in certain high-risk cases before 1 year of age, in cases of intermediate risk before 5 years of age, and in lower-risk patients later in childhood usually by 10 years of age. Because a limited number of mutations in *RET* cause more than 95% of hereditary MCT and up to 25% of sporadic disease, it is possible to screen the majority of patients using commercial reference laboratories.

Apparent sporadic MCT also calls for genetic testing and family studies. Up to 25% of new cases of MCT may actually be probands of families who harbor one of the familial syndromes. It has been noted in some kindreds, after careful and lengthy follow-up, that some individuals develop pheochromocytomas or primary hyperparathyroidism. Therefore, patients with

apparent familial MCT should be followed indefinitely for one of these components of MEN 2A. Screening of family members can be accomplished by genetic testing, once the index case is identified. Identification of a mutation that is present only in tumor tissue would establish the mutation as somatic and the tumor as a sporadic one. Identification of the same *RET* mutation in tumor and genomic DNA would make the diagnosis of a familial form of the disorder and would mandate careful screening of the family.

HYPERCALCEMIA

Clinical Features

A number of symptoms and signs accompany hypercalcemia. They include central nervous system effects such as lethargy, depression, psychosis, ataxia, stupor, and coma; neuromuscular effects such as weakness, proximal myopathy, and hypertonia; cardiovascular effects such as hypertension, bradycardia (and eventually asystole), and a shortened QT interval; renal effects such as stones, decreased glomerular filtration, polyuria, hyperchloremic acidosis, and nephrocalcinosis; gastrointestinal effects such as nausea, vomiting, constipation, and anorexia; eye findings such as band keratopathy; and systemic metastatic calcification. Primary hyperparathyroidism, one of the most common etiologies for hypercalcemia, has the mnemonic for recalling its signs and symptoms as *stones, bones, abdominal groans,* and *psychic moans* (Figure 8–12).

Stones
Renal stones
Nephrocalcinosis
Polyuria
Polydipsia
Uremia

Bones
Osteitis fibrosa with
-Subperiosteal resorption
-Osteoclastomas
-Bone cysts
Radiologic osteoporosis
Osteomalacia or rickets
Arthritis

1° HPTH signs and symptoms

Abdominal groans
Constipation
Indigestion, nausea, vomiting
Peptic ulcer
Pancreatitis

Psychic moans
Lethargy, fatigue
Depression
Memory loss
Psychoses-paranoia
Personality change, neuroses
Confusion, stupor, coma

Other
Proximal muscle weakness
Keratitis, conjunctivitis
Hypertension
Itching

FIGURE 8–12 Signs and symptoms of primary hyperparathyroidism.

TABLE 8–6 Causes of hypercalcemia.

Primary Hyperparathyroidism
Sporadic
Associated with MEN 1 or MEN 2A
Familial
Postrenal transplantation
Variant Forms of Hyperparathyroidism
Familial hypocalciuric hypercalcemia
Lithium therapy
Tertiary hyperparathyroidism in chronic renal failure
Malignancies
Humoral hypercalcemia of malignancy
Caused by PTHrP (solid tumors, adult T-cell leukemia syndrome)
Caused by 1,25(OH)$_2$D (lymphomas)
Caused by ectopic secretion of PTH (rare)
Local osteolytic hypercalcemia (multiple myeloma, leukemia, lymphoma)
Sarcoidosis or Other Granulomatous Diseases
Endocrinopathies
Thyrotoxicosis
Adrenal insufficiency
Pheochromocytoma
VIPoma
Drug Induced
Vitamin A intoxication
Vitamin D intoxication
Thiazide diuretics
Lithium
Milk-alkali syndrome
Estrogens, androgens, tamoxifen (in breast carcinoma)
Immobilization
Acute Renal Failure
Idiopathic Hypercalcemia of Infancy
ICU Hypercalcemia
Serum Protein Disorders

Mechanisms

Although many disorders are associated with hypercalcemia (Table 8–6), they produce it by a limited number of mechanisms: (1) increased bone resorption, (2) increased gastrointestinal absorption of calcium, or (3) decreased renal excretion of calcium. Although any of these mechanisms can be involved in a given patient, the common feature of virtually all hypercalcemic disorders is accelerated bone resorption. The only hypercalcemic disorder in which bone resorption does not play a part is the milk-alkali syndrome.

The central feature of the defense against hypercalcemia is suppression of PTH secretion. This reduces bone resorption, renal production of 1,25(OH)$_2$D, and, thereby, intestinal calcium absorption and increases urinary calcium losses. The kidney plays a key role in the adaptive response to hypercalcemia as the only route of net calcium elimination. The level of renal calcium excretion is markedly increased by the combined effects of an increased filtered load of calcium and the suppression of PTH. However, the patient who relies on the kidneys to excrete an increased calcium load is in precarious balance. Glomerular filtration is impaired by
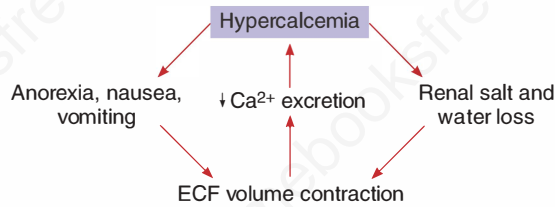
FIGURE 8–13 Vicious cycle of hypercalcemia. Once established, hypercalcemia can be maintained or aggravated as depicted. (Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism*, 3rd ed. New York: McGraw-Hill Education; 1995.)

hypercalcemia; the urinary concentrating ability is diminished, predisposing to dehydration; poor mentation may interfere with access to fluids; and nausea and vomiting may further predispose to dehydration and renal azotemia. Renal insufficiency, in turn, compromises calcium clearance, leading to a downward spiral (Figure 8–13). Thus, once established, many hypercalcemic states are self-perpetuating or aggravated through the *vicious cycle of hypercalcemia*. The only alternative to the renal route for

elimination of calcium from the extracellular fluid is deposition of calcium phosphate and other salts in bone and soft tissues. Soft tissue calcification is observed with massive calcium loads, with massive phosphate loads (as in tumor lysis syndrome, crush injuries and compartment syndromes), and when renal function is markedly impaired and the calcium-phosphate product rises.

Differential Diagnosis

As a practical matter, the categories for differential diagnosis are primary hyperparathyroidism and *everything else* (see Table 8–6). Hyperparathyroidism is by far the most common cause of outpatient hypercalcemia and has distinctive pathophysiologic features. Thus, the first step in the differential diagnosis is determination of intact PTH (Figure 8–14). If the PTH level is high, and thus inappropriate for hypercalcemia, little further workup is required except to consider the variant forms of hyperparathyroidism that are discussed later. If the PTH level is suppressed, then a search for other entities must be conducted. Most other entities in Table 8–6 are readily diagnosed by their distinctive features, as discussed later.
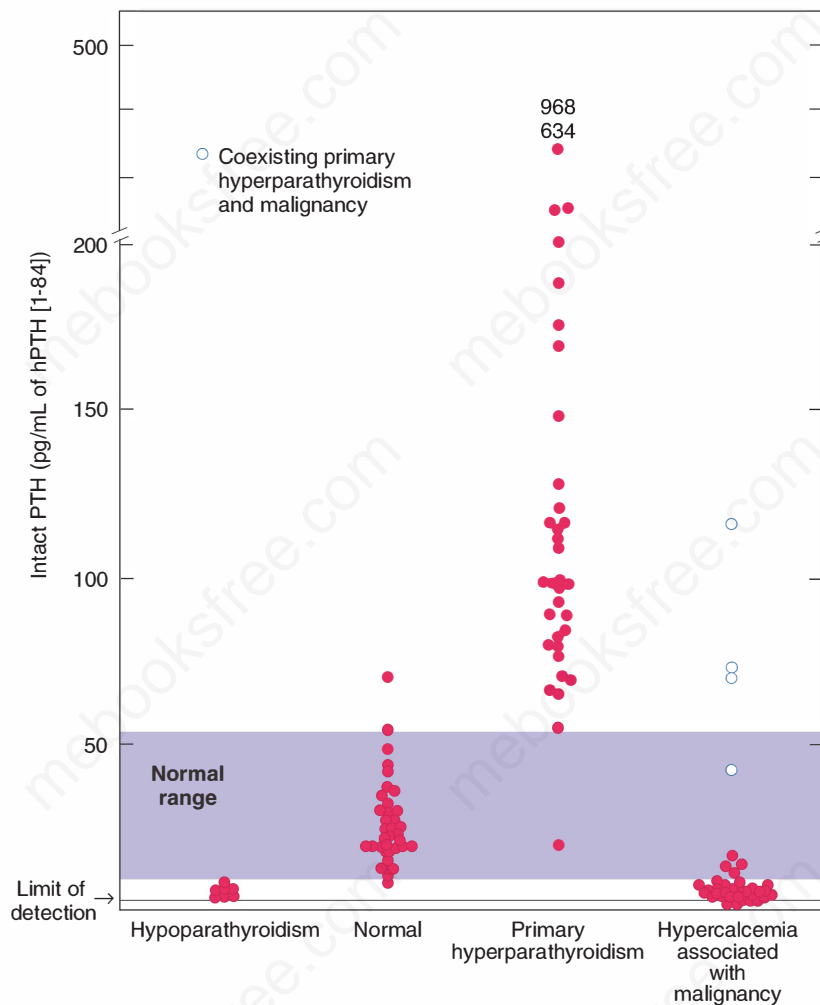
FIGURE 8–14 Clinical utility of immunoradiometric assay for intact PTH. (Reproduced with permission from Endres DB, Villanueva R, Sharp CF, et al. Measurement of parathyroid hormone. *Endocrinol Metab Clin North Am*. 1989 Sep;18(3):611-629.)

DISORDERS CAUSING HYPERCALCEMIA

1. PRIMARY HYPERPARATHYROIDISM

Primary hyperparathyroidism results from the excessive secretion of PTH and typically produces frank hypercalcemia. With routine screening of serum chemistries, we have come to recognize that primary hyperparathyroidism is a common and usually asymptomatic disorder. Its incidence is approximately 42 per 100,000, and its prevalence is up to 4 per 1000 in women over age 60. Primary hyperparathyroidism is approximately 2 to 3 times more common in women than in men.

Etiology and Pathogenesis

Primary hyperparathyroidism is caused by a single parathyroid adenoma in about 80% to 85% of cases and by primary hyperplasia of the parathyroids in 10% to 15%. Parathyroid carcinoma is a rare cause of hyperparathyroidism, accounting for 1% to 2% of cases. Parathyroid carcinoma is often recognizable preoperatively because it presents with severe hypercalcemia or a palpable neck mass.

Sporadic parathyroid adenomas have a clonal origin, indicating that they can be traced back to an oncogenic mutation in a single progenitor cell. A few of these genetic alterations have been identified and/or assigned to chromosomal loci. About 12% to 35% of sporadic parathyroid adenomas have inactivating mutations in the tumor suppressor gene *MENIN* (11q12-13). As reviewed later and in Chapter 22, loss of menin function is the cause of parathyroid, pituitary, and pancreatic tumors in the MEN 1 syndrome. An additional 40% of parathyroid adenomas display allelic loss on chromosome 1p (1p32pter).

Another important locus (11q13) has been implicated in approximately 8% of sporadic parathyroid adenomas and involves the *cyclin D1* gene. In the initial elucidation of this pathogenic mechanism, a chromosomal rearrangement was found in a parathyroid tumor. Breakage and then inversion of a piece of chromosome 11 led to the expression of cyclin D1, a cell-cycle regulatory protein, under the control of the PTH promoter. As expected, this promoter is highly active in parathyroid cells, resulting in marked overexpression of cyclin D1 in these tumor cells. Cyclin D1 is normally expressed at high levels in the G_1 phase of the cell cycle and permits entry of cells into the mitotic phase of the cycle. Thus, a parathyroid-specific disorder of cell-cycle regulation leads to abnormal cell proliferation and ultimately excessive PTH production.

Parathyroid hyperplasia accounts for 12% to 15% of cases of primary hyperparathyroidism and is the pathologic basis for these inherited conditions: MEN 1, MEN 2A, the hyperparathyroidism-jaw tumor (HPT-JT) syndrome, and isolated familial hyperparathyroidism. In all of these instances four or fewer glands may be involved at the time of presentation, or the hyperparathyroidism can occur in different glands asynchronously. Inheritance of all of these conditions is autosomal dominant.

Parathyroid hyperplasia appears to share similar pathogenic mechanisms with parathyroid adenomas. The MEN 1 syndrome illustrates this. MEN 1 is due to the inherited inactivation of one allele of *MENIN*, a tumor suppressor. Acquired postnatal somatic mutations in the remaining *MENIN* allele result in loss of that allele's function as well. This leads to tumors in those endocrine tissues in which the gene is expressed. In this view, multicentric somatic mutations would account for the occurrence of 4-gland hyperparathyroidism.

Menin is a key cellular protein responsible for transcriptional control and regulation of the cell cycle and other pathways important in cell growth. Over 1000 mutations have been uncovered in *MENIN* accounting for cases of MEN 1. Approximately 10% of MEN 1 cases are due to new mutations.

Kindreds with a presentation like MEN 1 but without mutations in *MENIN* have been investigated for germline mutations in cyclin-dependent kinase inhibitors (CDKIs) of which there are seven known genes. In an analysis of 196 consecutive cases of MEN 1 or other tumor states without germline mutations in *MENIN*, one laboratory found seven probable pathologic mutations in these genes. Three of the seven mutations were in *p27* with the remainder in three other CDKIs (*p15, p18,* and *p27*). These mutations are hypothesized to be disease-causing based on *in vitro* studies. The altered sequences in the CDKI protein can alter the downstream interactions they have with other proteins (such as the adaptor protein GRB2) or can alter protein stability and/or degradation. The CDKIs negatively control the cell cycle through interaction with and inhibition of cyclins D, E, and A. Germline mutations in kindreds would, therefore, involve loss of function in known inhibitors of cell growth and cell cycle activity.

In MEN 2A, parathyroid hyperplasia can occur due to activating mutations of *RET* in the four parathyroid glands. A *RET* mutation at codon 634 is one that predisposes to the development of primary hyperparathyroidism. Only 10% to 30% of patients with MEN 2A develop primary hyperparathyroidism.

In addition to classical MEN 1, germline mutations in *MENIN* have been identified in a subset of cases of familial isolated hyperparathyroidism. This observation established the concept that isolated familial hyperparathyroidism is an allelic variant of MEN 1 missing the tumors in the other endocrine glands. Other cases of familial isolated hyperparathyroidism are explained, albeit rarely, by mutations in *CDC73* the gene that causes the HPT-JT syndrome (discussed later). Additional families with isolated hyperparathyroidism have, as yet, no identified mutations.

In contrast to MEN 1 and MEN 2A, both benign and malignant parathyroid tumors occur in the HPT-JT syndrome. This is an autosomal dominant disorder that includes primary hyperparathyroidism (90%), jaw tumors (ossifying fibromas of the mandible or maxilla) (30%), renal cysts (10%), and less frequently renal hamartomas and Wilms tumor. Linkage analysis originally determined the locus for this syndrome, which had been designated as *hyperparathyroidism 2 or HRPT2* to be at 1q24-q32. Inactivating mutations in the tumor suppressor *CDC73*, encoding the protein parafibromin, have been identified in HPT-JT kindreds. These mutations inactivate parafibromin, a protein that plays a role in transcriptional regulation and chromatin modification.

In the HPT-JT syndrome, both single- and multiple-gland disease have been noted. It was further noted that approximately 15% of affected individuals with HPT-JT syndrome develop parathyroid cancer, instead of adenomas or hyperplasia, suggesting that reduced parafibromin function also contributes to a malignant phenotype. Several groups have shown that a large percentage (~70%) of patients with sporadic parathyroid cancer harbor somatic mutations in *CDC73*. Surprisingly, approximately 30% of individuals with isolated parathyroid cancer have also been found to have germline mutations in *CDC73* without the other manifestations of the HPT-JT syndrome. It is now thought that the majority of sporadically occurring parathyroid cancers are due to somatic mutations in the coding region of this gene. A pathologic study of more than 50 cases of parathyroid cancer indicates that the loss of parafibromin immunoreactivity is a specific and key feature of malignancy with a sensitivity of 96% (CI, 85%-99%) and specificity of 99% (CI, 92%-100%).

Clinical Features

A. Symptoms and signs

The typical clinical presentation of primary hyperparathyroidism has evolved considerably over the past few decades. As the disease continues to be detected by screening that includes determination of serum calcium levels, there has been a marked reduction in the frequency of the classic signs and symptoms of primary hyperparathyroidism. Renal disease (stones, decreased renal function, and occasionally nephrocalcinosis) and the classic hyperparathyroid bone disease osteitis fibrosa cystica are decidedly rare today. In fact, about 85% of patients presenting today have neither bone nor renal manifestations of hyperparathyroidism and are regarded as asymptomatic or at most minimally symptomatic. At the same time, we have begun to recognize more subtle manifestations of hyperparathyroidism. This has presented a number of questions about the role of parathyroid surgery in primary hyperparathyroidism, which are discussed later (see Treatment).

1. **Hyperparathyroid bone disease**—The classic bone disease of hyperparathyroidism is osteitis fibrosa cystica. Formerly common, this disorder now occurs in less than 10% of patients. Clinically, osteitis fibrosa cystica causes bone pain and sometimes pathologic fractures. The most common laboratory finding is an elevation of the alkaline phosphatase level, reflecting high bone turnover. Histologically, there is an increase in the number of bone-resorbing osteoclasts, marrow fibrosis, and cystic lesions that may contain fibrous tissue (brown tumors). The most sensitive and specific radiologic finding of osteitis fibrosa cystica is subperiosteal resorption of cortical bone, best seen in high-resolution films of the phalanges (Figure 8–15A). A similar process in the skull leads to a salt-and-pepper appearance (Figure 8–15B). Bone cysts or brown tumors may present as osteolytic lesions. Dental films may disclose loss of the lamina dura of the teeth, but this is a nonspecific finding also seen in periodontal disease.

    The other important skeletal consequence of hyperparathyroidism is osteoporosis or low bone mass. Unlike other osteoporotic disorders, hyperparathyroidism often results in the preferential loss of cortical bone (Figure 8–16). In general, both the mass and mechanical strength of trabecular bone are relatively maintained in mild primary hyperparathyroidism. Patients who are followed medically with this disease generally do not experience progressive bone loss for as long as 8 years after diagnosis (Figure 8–17). This may be due to the fact that mild PTH excess has an anabolic effect on the skeleton to maintain or even increase bone mass. However, longer medical
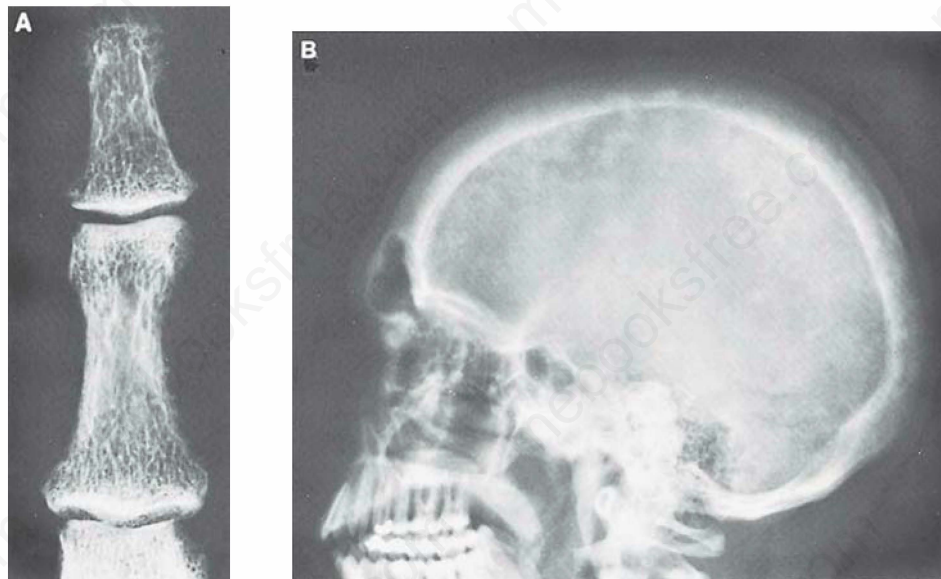
FIGURE 8–15   A. Magnified x-ray of index finger on fine-grain industrial film showing classic subperiosteal resorption in a patient with severe primary hyperparathyroidism. Note the left (radial) surface of the distal phalanx, where the cortex is almost completely resorbed, leaving only fine wisps of cortical bone. **B.** Skull x-ray from a patient with severe secondary hyperparathyroidism due to end-stage renal disease. Extensive areas of demineralization alternate with areas of increased bone density, resulting in the "salt and pepper" skull x-ray. (Used with permission from Dr. Harry Genant.)
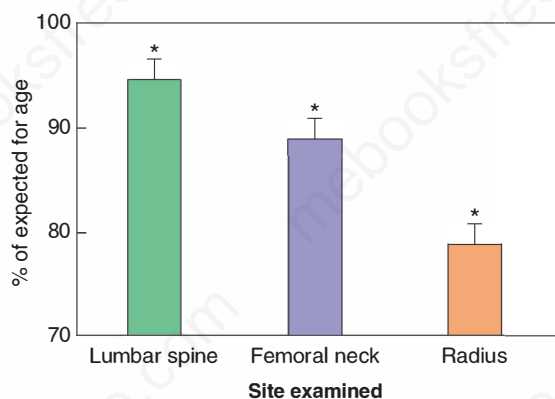
FIGURE 8–16 Bone mineral density at several sites in primary hyperparathyroidism shown as the percentage of expected for age. (Adapted with permission from Silverberg SJ, et al. Nephrolithiasis and bone involvement in primary hyperparathyroidism. *Am J Med.* 1990 Sep;89(3):327-334.)

follow-up of patients with mild primary hyperparathyroidism (10-15 years) demonstrates that bone mineral density (BMD) by dual energy x-ray absorptiometry (DXA) declines significantly at the hip and forearm. Osteoporosis, defined by BMD T-scores by DXA (–2.5 or lower) is generally considered to be an indication for surgical treatment of primary hyperparathyroidism (see Treatment).

2. **Hyperparathyroid kidney disease**—Once common in primary hyperparathyroidism, kidney stones now occur in 10% to 20% of cases depending on the series. These are usually calcium oxalate stones. From the perspective of a stone clinic, only about 7% of calcium stone-formers prove to have primary hyperparathyroidism. They are difficult to manage medically, and stones constitute one of the agreed indications for parathyroidectomy. Surgical cure of primary hyperparathyroidism in patients who form stones produces a substantial reduction in future stone formation. Clinically evident nephrocalcinosis rarely occurs, but a gradual loss of renal function is not uncommon. Renal function is stabilized after a successful parathyroidectomy, and otherwise unexplained renal insufficiency (defined as an estimated glomerular filtration rate [eGFR] <60 mL/min) in the setting of primary hyperparathyroidism is also considered to be an indication for surgery because of the risk of progression. Chronic hypercalcemia can also compromise the renal concentrating ability, giving rise to polydipsia and polyuria and dehydration.

3. **Nonspecific features of primary hyperparathyroidism**—Although stupor and coma occur in severe hypercalcemia, the degree to which milder impairments of central nervous system function affect the typical patient with primary hyperparathyroidism is unclear. Lethargy, fatigue, depression, difficulty in concentrating, and personality changes may occur; however, some patients appear to benefit from parathyroidectomy. Frank psychosis also responds to surgery on occasion. Muscle weakness with characteristic electromyographic changes is also seen, and there is evidence from controlled clinical trials that surgery can improve muscle strength. A number of studies have examined quality of life (QOL) and psychological function in patients with primary hyperparathyroidism pre- and post-parathyroid surgery, compared to a medically followed group. Results vary as to whether QOL improves with surgery. It was

formerly thought that the incidence of hypertension was increased in primary hyperparathyroidism, but that condition appears not to be more common compared to age-matched controls, and parathyroidectomy appears to be of no benefit in curing hypertension. Dyspepsia, nausea, and constipation all occur, probably as a consequence of hypercalcemia, but there is probably no increase in the incidence of peptic ulcer disease. The articular manifestations of primary hyperparathyroidism include chondrocalcinosis in up to 5% of patients. Acute attacks of pseudogout, however, are less frequent.

B. Laboratory findings   Hypercalcemia is nearly universal in primary hyperparathyroidism, although the serum calcium sometimes fluctuates into the upper normal range. In patients with subtle hyperparathyroidism, repeated serum calcium measurements over a period of time may be required to establish the pattern of intermittent hypercalcemia. Both total and ionized calcium are elevated, and in most clinical instances there is no advantage to measuring the ionized calcium level. Patients with normocalcemic variant of primary hyperparathyroidism, in whom subtle vitamin D deficiency and malabsorption have been eliminated, are being recognized more frequently. In patients with primary hyperparathyroidism, the serum phosphorus level is low-normal (<3.5 mg/dL) or low (<2.5 mg/dL) because of the phosphaturic effect of PTH. A mild hyperchloremic metabolic acidosis may also manifest itself.

The diagnosis of primary hyperparathyroidism in a hypercalcemic patient can be made by determining the intact PTH level. As shown in Figure 8–14, an elevated or even upper-normal level of PTH is clearly inappropriate in a hypercalcemic patient and establishes the diagnosis of hyperparathyroidism or one of its variants—FHH or lithium-induced hypercalcemia. The reliability of the two-site intact PTH assay allows for the definitive diagnosis of primary hyperparathyroidism. In a patient with a high PTH level, there is no need to screen for such disorders as metastatic cancer or sarcoidosis. Determinations of BMD, renal function, and renal imaging [abdominal radiograph, ultrasound, computed tomography (CT) scan] for renal stones are often obtained for clinical decision-making reasons. A determination of the ratio of calcium clearance and creatinine clearance should always be obtained to help to exclude FHH from consideration, particularly in the asymptomatic patient.

Treatment

The definitive treatment of primary hyperparathyroidism is parathyroidectomy. The surgical strategy (ie, minimally invasive vs bilateral neck exploration) depends on the ability of localizing studies such as sestamibi scanning and cervical ultrasound to identify one clearly abnormal gland and the availability of intraoperative PTH determinations to verify that the disease-producing lesion has been removed during surgery. If multiple enlarged glands are suspected, the likely diagnosis is parathyroid hyperplasia or double adenoma. In patients with hyperplasia, the preferred operation is a 3½ gland parathyroidectomy, leaving a remnant sufficient to prevent hypocalcemia. Double parathyroid adenomas are both removed in affected patients. The pathologist is of little
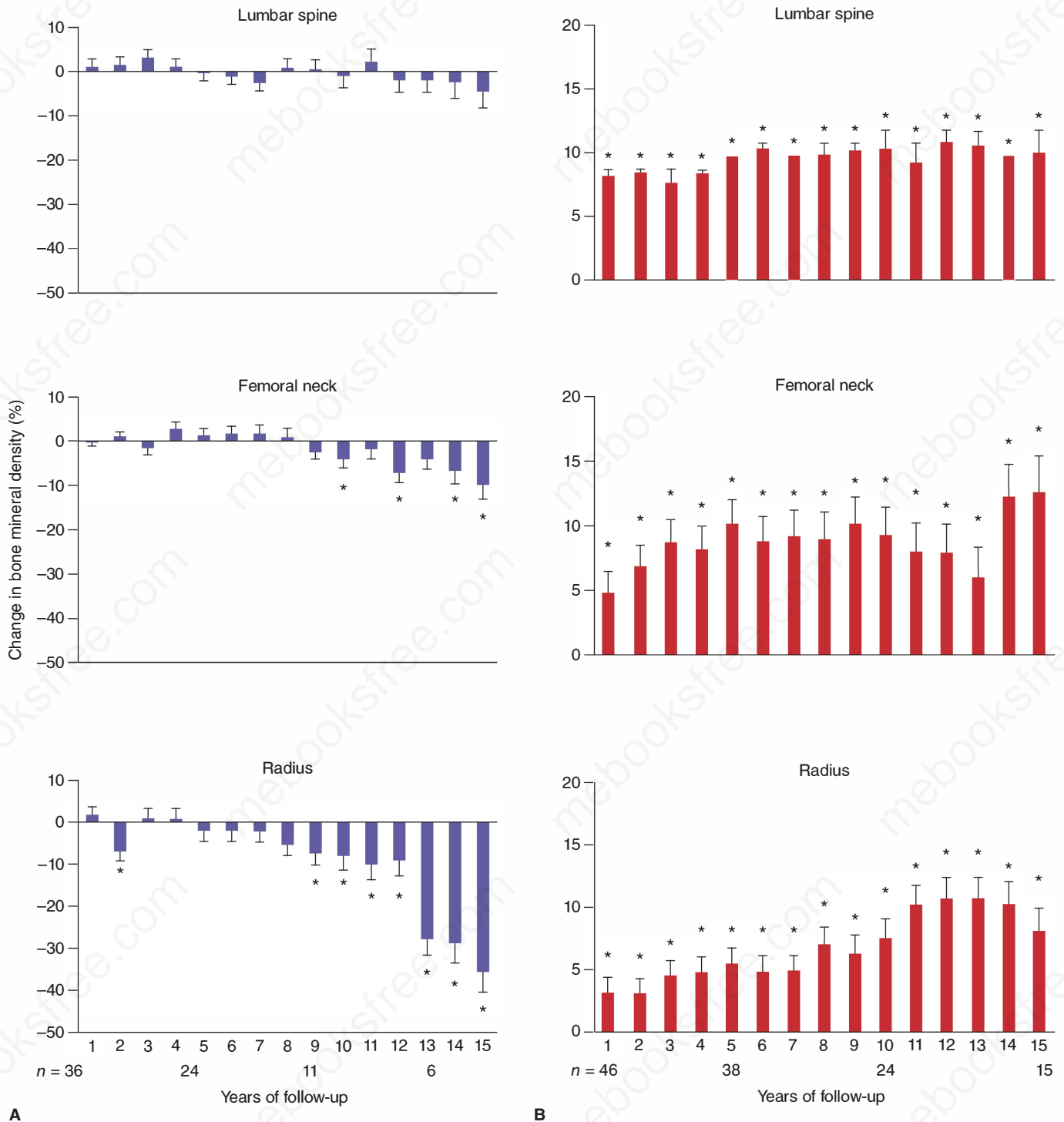
FIGURE 8–17 Mean changes in bone mineral density (BMD) measurements by dual-energy x-ray absorptiometry (DXA) in patients who underwent parathyroidectomy compared with medical follow-up during a 15-year observational study. BMD changes are reported compared with baseline and were statistically different from baseline (*$P < 0.05$). The number of patients whose measurements are included is shown underneath each time point. Either just before (distal 1/3 radius) or after 10 years of conservative follow-up (femoral neck) BMD measurements began to fall significantly, while lumbar spine BMD remained unchanged (**Panel A**). Patients who underwent parathyroidectomy demonstrated improved BMD which persisted during 15 years of follow-up (**Panel B**). (Reproduced with permission from Rubin MR[1], Bilezikian JP, McMahon DJ, et al. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years. *J Clin Endocrinol Metab*. 2008 Sep;93(9):3462-3470.)

help in distinguishing among adenoma and hyperplasia; both typically have a robust degree of cellularity. However, the percentage of fat cells in the gland(s) removed, especially if substantial suggests the gland is a normal one. The recurrence rate of hypercalcemia is high in patients who have parathyroid hyperplasia—particularly in those with MEN 1 or allelic variants of MEN 1 or the HPT-JT syndromes, because of the inherited propensity for tumor growth. In such cases, the parathyroid remnant can be removed from the neck and implanted in pieces in forearm muscles to allow for easy subsequent removal of some parathyroid tissue if hypercalcemia recurs.

In experienced hands, the cure rate for a single parathyroid adenoma is more than 95%. The success rate in primary parathyroid hyperplasia is somewhat lower, because of missed glands and recurrent hyperparathyroidism in patients with MEN syndromes. There is a 20% incidence of persistent or recurrent hypercalcemia. Parathyroidectomy in MEN 1 patients can be difficult. Normal parathyroid glands weigh only about 40 mg and may be located throughout the neck or upper mediastinum. It is mandatory not only to locate the abnormal parathyroid gland(s), but also to find the other gland or glands and determine whether or not they are normal in size. Complications of parathyroid surgery in general include damage to the recurrent laryngeal nerve, which passes close to the posterior thyroid capsule, and inadvertent removal or devitalization of all parathyroid tissue, producing permanent hypoparathyroidism. In skilled hands, the incidence of these complications is less than 1%. It is critical that parathyroid surgery be performed by a surgeon with specialized skill and experience (see Chapter 26).

Parathyroid gland localization studies and intraoperative PTH testing are critical components of the surgical management of patients with primary hyperparathyroidism. If localizing studies clearly indicate a single abnormal gland and if intraoperative PTH testing is available, surgical management generally consists of unilateral exploration with the removal of the single abnormal gland. Localization studies continue to be essential in the management of patients with recurrent or persistent hyperparathyroidism. The most successful procedures are ^{99m}Tc-sestamibi scanning, CT scanning (including 4-D CT), magnetic resonance imaging, and ultrasound. Individually, each has a sensitivity of 60% to 80% in experienced hands. Used in combination, they are successful in at least 80% of reoperated cases. Invasive studies, such as angiography and venous sampling, are reserved for the most difficult cases.

There is no definitive medical therapy for primary hyperparathyroidism. In postmenopausal women, estrogen therapy in high doses (1.25 mg of conjugated estrogens or 30-50 μg of ethinyl estradiol) produces an average decrease of 0.5 to 1 mg/dL in the serum calcium and an increase in BMD. The effects of estrogen are on the skeletal responses to PTH. PTH levels do not fall. Estrogen therapy is almost never used for this condition due to the potential for complications in postmenopausal women. Small clinical trials provide experience with oral bisphosphonates (eg, alendronate) and the selective estrogen response modulator (SERM) raloxifene in these patients. Treatment with alendronate for 1 or 2 years reduces biochemical markers of bone turnover and improves BMD by DXA, especially in the spine and also in the femoral neck,

compared to baseline values prior to treatment. Studies have shown that raloxifene, administered over several weeks, to postmenopausal women with mild primary hyperparathyroidism also reduces bone turnover markers significantly. Neither alendronate nor raloxifene have significant and/or persistent effects on serum calcium or intact PTH levels, nor is either agent approved for treatment of primary hyperparathyroidism by the U.S. Food and Drug Administration (FDA). Calcimimetic agents that activate the parathyroid CaSR offer an alternative to surgery for individuals who meet guidelines for surgical referral but who are too sick or unwilling to undergo surgery or in whom surgery has failed to cure their hyperparathyroidism. The calcimimetic cinacalcet is FDA-approved for the management of primary hyperparathyroidism in the United States with those indications and for the treatment of secondary hyperparathyroidism in patients with CKD on dialysis (discussed later). In small clinical trials, cinacalcet has been shown to normalize serum [Ca^{2+}] and raise serum phosphate with modest lowering of PTH levels in patients with mild asymptomatic primary hyperparathyroidism for up to 5 years of therapy. No significant changes in BMD by DXA were seen. Patients with moderate to severe primary hyperparathyroidism (serum [Ca^{2+}] >12.5 mg/dL) also responded to cinacalcet with lowering of serum [Ca^{2+}] from 12.7 ± 0.8 mg/dL at baseline by 1 mg/dL or greater in 88% of patients. QOL parameters also improved compared to baseline. Therapy targeted to the CaSR is an alternative for controlling hypercalcemia in these patients.

The relatively asymptomatic status of most patients with primary hyperparathyroidism today presents a dilemma. Which of them should be subjected to surgery? More information on the natural history of untreated primary hyperparathyroidism would help in resolving this uncertainty. Most observational studies lack long-term follow-up and appropriate control groups. However, one observational study of 15 years duration showed modest increases in serum [Ca^{2+}] (after 10 years) and slow but consistent downward trends in BMD at cortical sites (femoral neck and distal radius) (after 8-10 years of follow-up) (see Figure 8–17A). This study has challenged the skeletal safety of long-term observation of otherwise asymptomatic patients. On the other hand, surgery is usually curative. In experienced hands, surgery has a low morbidity rate. Although parathyroid surgery has a substantial initial cost, over the long term, the cost-benefit ratio may be favorable when compared with a lifetime or many years of medical follow-up. Moreover, there is a marked improvement in BMD after surgery (see Figure 8–17B), with sustained increases over 15 years postoperatively.

Consensus Development Conferences or International Workshops in 1990, 2002, 2008, and 2013 examined evidence and published recommendations for surgery in mild asymptomatic primary hyperparathyroidism and for conservative management. Current recommendations for surgery (2013) are: (1) serum [Ca^{2+}] greater than 1 mg/dL above the upper limits of normal; (2) BMD DXA T score less than –2.5 at the lumbar spine, total hip, femoral neck or one-third distal radius; (3) presence of vertebral fracture(s) by any imaging modality; (4) estimated glomerular filtration rate (eGFR) less than 60 mL/min; (5) 24 hour urine calcium excretion greater than 400 mg and increased stone risk by biochemical stone

risk analysis; (6) presence of nephrolithiasis or nephrocalcinosis by x-ray, ultrasound, or CT; (7) age less than 50 years. Patients with a coexisting illness complicating the management of hypercalcemia should also be considered for surgery.

Recommendations for medical monitoring of patients not undergoing surgery should include annual measurements of serum [Ca^{2+}], PTH, creatinine, and eGFR. If renal stones have occurred or are suspected, biochemical analyses of the urine and renal imaging should be done. Skeletal surveillance should include DXA scanning every 1 or 2 years or as clinically appropriate and/ or spinal x-rays.

Variants of Primary Hyperparathyroidism

A. Familial hypocalciuric hypercalcemia   Inherited as an autosomal dominant trait, this disorder is responsible for life-long generally asymptomatic hypercalcemia, first detectable in cord blood. Hypercalcemia is usually mild (10.5-12 mg/dL [2.7-3.0 mmol/L]) and is often accompanied by mild hypophosphatemia and hypermagnesemia. The PTH level is normal or slightly elevated, indicating that this is a PTH-dependent form of hypercalcemia. The parathyroid glands are normal in size or slightly enlarged. The most notable laboratory feature of the disorder is hypocalciuria, which nearly all patients demonstrate. The urinary calcium level is usually less than 50 mg/24 h, and the calcium/creatinine clearance ratio is usually less than 0.01 and calculated as follows:

$$\frac{\text{Urine calcium} \times \text{Serum creatinine}}{\text{Serum calcium} \times \text{Urine creatinine}}$$

Hypocalciuria is an intrinsic renal trait; it persists even in patients who have undergone total parathyroidectomy and become hypocalcemic.

Because FHH is typically asymptomatic and benign, the most important role of diagnosis is to distinguish it from primary hyperparathyroidism and avoid an unnecessary parathyroidectomy. If subtotal parathyroidectomy is performed, the serum calcium invariably returns shortly to preoperative levels; patients with FHH resist attempts to lower the serum calcium while functioning parathyroid tissue remains. Unfortunate patients with this condition who have undergone total parathyroidectomy are rendered hypoparathyroid and are dependent on treatment with calcium and vitamin D supplements.

The diagnosis must be considered in patients with asymptomatic mild hypercalcemia who are relatively hypocalciuric. However, an unequivocal diagnosis cannot be made biochemically because the serum and urinary calcium and PTH levels all overlap with typical primary hyperparathyroidism. Family screening and/ or genetic testing are part of the workup needed to confirm the diagnosis. The penetrance of the phenotype is essentially 100%, and affected family members are hypercalcemic throughout life, so if the proband has the disorder, each first-degree relative who is screened has a 50% chance of being hypercalcemic.

The majority of cases of FHH (60%-70%) are caused by loss-of-function mutations in the CaSR, and this is termed FHH type 1.

Loss of the functional output of one CaSR allele shifts the set-point for inhibition of PTH release to the right, producing hypercalcemia. The same receptor is expressed in the kidney, where it regulates renal calcium excretion. Recently, patients with the FHH phenotype (hypercalcemia and hypocalciuria) who did not harbor mutations in the *CaSR* were shown to have mutations in *G-alpha 11* (FHH type 2) and a subunit of an adapter protein complex 2σ1 subunit (*AP2S1*; FHH type 3). A large number of point mutations in different exons of the CaSR produce the FHH type 1 phenotype. Fewer mutations and probands with FHH 2 and 3 have been so far identified. However, these genetic disorders must be considered in the full molecular evaluation of a proband and family with the FHH phenotype, especially those who have demonstrated no response in serum calcium to parathyroidectomy. Acquired hypocalciuric hypercalcemia has been described in association with antibodies that interact with the CaSR and block the ability of high [Ca^{2+}] to suppress PTH.

A child of two parents with FHH type 1 may inherit a mutant allele from each parent, producing **neonatal severe hypercalcemia**, a life-threatening disorder in which failure to sense extracellular calcium causes severe hyperparathyroidism, requiring total parathyroidectomy soon after birth. Clinically significant hypercalcemia has also been seen in babies with heterozygous *CaSR* mutations, if the mutation produces a dramatic loss of function in the CaSR protein.

B. MEN syndromes   As noted earlier, primary hyperparathyroidism is a feature of both MEN 1 and MEN 2A (see also Chapter 22). The penetrance of primary hyperparathyroidism in MEN 1 is over 90% by age 40. Patients with the MEN 1 syndrome inherit a loss-of-function germline mutation in the tumor suppressor gene *MENIN* on chromosome 11q12-13. A mutation during mitosis of a parathyroid cell that resulted in loss of function in the remaining allele in that cell would abrogate the cell's growth control mechanism and permit clonal expansion of its progeny ultimately to generate the parathyroid tumor. A similar mechanism appears also to operate in a small fraction of sporadic parathyroid adenomas with 11q12-13 deletions.

The penetrance of primary hyperparathyroidism in MEN 2A is up to 35%. As discussed in the earlier section on MCT, the disorder is caused by activating mutations in *RET*, a gene encoding tyrosine kinase growth factor receptor. Evidently, the *RET* gene product is less important for growth of the parathyroids than for thyroid C cells, because the penetrance of primary hyperparathyroidism is fairly low and because (in MEN 2B) a separate class of activating *RET* mutations produces MCT and pheochromocytoma and very rarely primary hyperparathyroidism. The treatment for parathyroid hyperplasia in MEN 1 or MEN 2 is usually subtotal parathyroidectomy. The recurrence rate is higher than in sporadic parathyroid hyperplasia and may approach 50% in MEN 1.

C. Lithium therapy   Both in patients and in isolated parathyroid cells, exposure to extracellular lithium shifts the

set-point for inhibition of PTH secretion to the right. Clinically, this results in hypercalcemia and a detectable or elevated level of PTH. Lithium treatment can also produce hypocalciuria and thus can phenocopy FHH. Most patients with therapeutic lithium levels for bipolar affective disorder have a slight increase in the serum calcium level, and up to 10% become mildly hypercalcemic, with PTH levels that are high normal or slightly elevated. Lithium treatment can also unmask underlying primary hyperparathyroidism. It is difficult to diagnose primary hyperparathyroidism in a lithium-treated patient, particularly when temporary cessation of lithium therapy is difficult. However, the likelihood of underlying primary hyperparathyroidism is high when the serum calcium is greater than 11.5 mg/dL, and the decision to undertake surgery must be based on clinical criteria. Unfortunately, surgical cure of hyperparathyroidism rarely ameliorates the underlying psychiatric condition.

2. MALIGNANCY-ASSOCIATED HYPERCALCEMIA

Malignancy-associated hypercalcemia is the second most common form of hypercalcemia, with an incidence of 15 cases per 100,000 per year—about one-half the incidence of primary hyperparathyroidism. Approximately 20% of patients with cancer develop hypercalcemia, which is due to humoral mechanisms in 80% of patients and osseous metastases in the remaining 20%. Malignancy-associated hypercalcemia is much less prevalent than primary hyperparathyroidism, because most patients have such limited survival once this complication occurs. Nonetheless, malignancy-associated hypercalcemia is the most common cause of hypercalcemia in hospitalized patients. The clinical features and pathogenesis of malignancy-associated hypercalcemia are presented in Chapter 21. The treatment of nonparathyroid hypercalcemia is presented later.

3. SARCOIDOSIS AND OTHER GRANULOMATOUS DISORDERS

Hypercalcemia is seen in up to 10% of subjects with sarcoidosis. A higher percentage of patients with sarcoidosis have hypercalciuria, and that manifestation often predates the presentation of hypercalcemia. This is due to inappropriately elevated $1,25(OH)_2D$ levels and abnormal vitamin D metabolism. Lymphoid tissue and pulmonary macrophages from affected individuals contain 25(OH)D 1-hydroxylase activity that is not seen in normal individuals. The 1-hydroxylase enzyme, responsible for the overproduction of $1,25(OH)_2D$ in sarcoidosis, is the same as that in the kidney (CYP27B1). 1-Hydroxylase activity in these cells is not readily inhibited by calcium or $1,25(OH)_2D$, indicating a lack of feedback inhibition. Macrophages expressing the 1-hydroxylase in these patients do not express a normally functioning 24-hydroxylase so that the catabolism and clearance of $1,25(OH)_2D$ in these cells are lacking. This makes these subjects vulnerable to hypercalcemia or hypercalciuria during periods of increased vitamin D production (eg, summertime with increased sunlight exposure). γ-IFN stimulates 1-hydroxylase activity in these cells, which makes such subjects more vulnerable to altered calcium homeostasis when their disease is active. Glucocorticoids, on the other hand, suppress the inflammation, and subsequently 1-hydroxylase activity, as well as antagonize $1,25(OH)_2D$ action, providing effective treatment for both the disease and this complication of it.

Other granulomatous diseases are associated with abnormal vitamin D metabolism, resulting in hypercalcemia and/or hypercalciuria. These disorders include tuberculosis, berylliosis, disseminated coccidioidomycosis, histoplasmosis, leprosy, and pulmonary eosinophilic granulomatosis. Furthermore, a substantial number of subjects with Hodgkin or non-Hodgkin lymphomas develop hypercalcemia associated with inappropriately elevated $1,25(OH)_2D$ levels. In a recent series of 101 patients with elevated $1,25(OH)_2D$ levels and hypercalcemia, Donovan and colleagues found the leading causes to be sarcoidosis (49%), hematologic malignancies (17%), and infections (mainly mycobacterial, 8%). The median angiotensin-converting enzyme (ACE) levels in those with sarcoidosis were 218 U/L (interquartile range, 176-277; normal range 20-130 U/L). A level above the normal range did not reliably associate with sarcoidosis. Levels above 250 U/L, however, were specific (89%) but insensitive (31%). In addition, levels of $1,25(OH)_2D$ greater than 300 pmol/L (normal range, 40-150 pmol/L) were not observed in patients from this series with sarcoidosis. Although $1,25(OH)_2D$-mediated hypercalcemia is a rare occurrence, its etiology must be determined to select the most appropriate management.

4. ENDOCRINOPATHIES

Thyrotoxicosis

Mild hypercalcemia is found in about 10% of patients with thyrotoxicosis. The PTH level is suppressed, and the serum phosphate is in the upper normal range. The serum alkaline phosphatase and biochemical markers of bone turnover may be mildly increased. Significant hypercalcemia may develop in patients with thyrotoxicosis, particularly if it is severe and if patients are temporarily immobilized. Thyroid hormone has direct bone-resorbing properties causing a high turnover state, which often eventually progresses to mild osteoporosis.

Adrenal Insufficiency

Hypercalcemia can be a feature of acute adrenal crisis and responds rapidly to glucocorticoid therapy. Animal studies suggest that hemoconcentration is a critical factor. In experimental adrenal insufficiency, ionized $[Ca^{2+}]$ is normal.

5. ENDOCRINE TUMORS

Hypercalcemia in patients with pheochromocytoma may be a manifestation of primary hyperparathyroidism in MEN 2A. Hypercalcemia, however, is found occasionally in uncomplicated pheochromocytoma, paraganglioma, and other neuroendocrine

tumors where it appears to result from secretion of PTHrP by the tumor. About 40% of tumors secreting VIP (VIPomas) are also associated with hypercalcemia. The cause is unknown. It is known, however, that high levels of VIP may activate PTH/PTHrP receptors.

6. THIAZIDE DIURETICS

The administration of thiazides and related diuretics, such as chlorthalidone, metolazone, and indapamide, can produce an increase in the serum calcium that is not fully accounted for by hemoconcentration. Hypercalcemia is mild and usually transient, lasting for days or weeks, but occasionally it persists. Severe hypercalcemia is rare. In a retrospective series of patients with thiazide-induced hypercalcemia, 71% had persistent hypercalcemia after discontinuation of the drug. Primary hyperparathyroidism was diagnosed in 24% of those patients. Thiazide administration can also exacerbate the effects of underlying primary hyperparathyroidism.

7. VITAMIN D AND VITAMIN A

Hypervitaminosis D

Hypercalcemia may occur in individuals ingesting large doses of vitamin D either therapeutically or accidentally (eg, irregularities in milk product supplementation with vitamin D have been reported). The initial signs and symptoms of vitamin D intoxication include weakness, lethargy, headaches, nausea, and polyuria and are attributable to hypercalcemia and hypercalciuria. Ectopic calcification may occur, particularly in the kidneys, resulting in nephrolithiasis or nephrocalcinosis; other sites include blood vessels, heart, lungs, and skin. Infants appear to be quite susceptible to vitamin D intoxication and may develop disseminated atherosclerosis, supravalvular aortic stenosis, and renal acidosis.

Hypervitaminosis D is readily diagnosed by the very high serum levels of 25(OH)D, because the conversion of vitamin D to 25(OH)D is not tightly regulated. In contrast, $1,25(OH)_2D$ levels are often normal, but not suppressed. This reflects the expected feedback regulation of $1,25(OH)_2D$ production by the elevated calcium and reduced PTH levels. Levels of free $1,25(OH)_2D$ when measured have been found to be increased. This is in part caused by the high levels of 25(OH)D that displace $1,25(OH)_2D$ from DBP, raising the ratio of free:total $1,25(OH)_2D$. The elevated free concentration of $1,25(OH)_2D$, plus the intrinsic biologic effects of the elevated 25(OH)D concentration, combine to increase intestinal calcium absorption and bone resorption. The hypercalciuria, which is invariably seen, may lead to dehydration and coma as a result of hyposthenuria, prerenal azotemia, and worsening hypercalcemia due, in part, to decreased renal clearance.

The dose of vitamin D required to induce toxicity varies among patients, reflecting differences in absorption, storage, and subsequent metabolism of the vitamin as well as in target tissue response to the active metabolites. For example, an elderly patient is likely to have reduced intestinal calcium transport and renal production of $1,25(OH)_2D$. Such an individual may be able to tolerate 50,000 to 100,000 units of vitamin D daily (although these levels far exceed daily requirements). However, patients with unsuspected hyperparathyroidism receiving such doses for the treatment of osteoporosis are more likely to experience hypercalcemia. Treatment consists of withdrawing the vitamin D, rehydration, reducing calcium intake, and administering glucocorticoids, which antagonize the ability of $1,25(OH)_2D$ to stimulate intestinal calcium absorption. Excess vitamin D is slowly cleared from the body (weeks to months), so treatment is prolonged.

Hypervitaminosis A

Excessive ingestion of vitamin A, usually from self-medication with vitamin A preparations, causes a number of abnormalities, including gingivitis, cheilitis, erythema, desquamation, and hair loss. Bone resorption is increased, leading to osteoporosis and fractures, hypercalcemia, and hyperostosis. Excess vitamin A causes hepatosplenomegaly with hypertrophy of fat storage cells, fibrosis, and sclerosis of central veins. Many of these effects can be attributed to the effects of vitamin A on cellular membranes. Under normal circumstances, such effects are prevented because vitamin A is bound to retinol-binding protein (RBP), and its release from the liver is regulated. In vitamin A toxicity, however, these protective mechanisms are overcome, and retinol and its retinyl esters appear in blood unbound to RBP. The mechanism by which vitamin A stimulates bone resorption is not clear.

8. MILK-ALKALI SYNDROME

The ingestion of large quantities of calcium together with an absorbable alkali can produce hypercalcemia with alkalosis, renal impairment, and often nephrocalcinosis. The syndrome was more common when absorbable antacids were the standard treatment for peptic ulcer disease, but it is still seen occasionally. This is the only recognized example of pure absorptive hypercalcemia. The details of its pathogenesis are poorly understood.

9. MISCELLANEOUS CONDITIONS

Immobilization

In immobilized patients there is a marked increase in bone resorption, which often produces hypercalciuria and occasionally hypercalcemia, mainly in individuals with a preexisting high bone turnover state, such as adolescents and patients with thyrotoxicosis or Paget disease. Intact PTH and PTHrP levels are suppressed. The disorder remits with the restoration of activity. If acute treatment is required, bisphosphonates appear to be the treatment of choice.

Acute Renal Failure

Hypercalcemia is often seen when renal failure is precipitated by rhabdomyolysis and usually occurs during the early recovery stage, presumably as calcium deposits are mobilized from damaged muscle tissue. It typically resolves over a few weeks.

TREATMENT OF HYPERCALCEMIA

Initial management of hypercalcemia consists of assessing the hydration state of the patient and rehydrating as necessary with intravenous saline. The first goal is to restore renal function, which is often impaired in hypercalcemia because of reduced glomerular filtration and dehydration. Hypercalcemia impairs the urinary concentrating ability, leading to polyuria, and at the same time impairs the sensorium, diminishing the sense of thirst. Once renal function is restored, excretion of calcium can be further enhanced by inducing a saline diuresis. Because most of the filtered calcium is reabsorbed by bulk flow in the proximal tubule along with sodium chloride, a saline diuresis markedly increases calcium excretion. However, a vigorous saline diuresis also induces substantial urinary losses of potassium and magnesium, and these must be monitored and replaced as necessary.

In contemporary management, prolonged saline infusion has been supplanted by early institution, when safe, of intravenous bisphosphonate therapy. Chronic maintenance therapy may be started soon after hospitalization, once the diagnosis is secured, and adequate hydration and renal perfusion are established. The most useful agents for chronic hypercalcemia due to cancer are intravenous bisphosphonates (zoledronic acid or pamidronate), which can take up to 5 days to have their full effect. Bisphosphonates act by inhibiting osteoclastic bone resorption. The initial dose of pamidronate is 60 to 90 mg by intravenous infusion over 1 hour, and the dose of zoledronic acid is 4 mg infused over 15 minutes. In two large trials, 88% and 70% of patients with malignancy-associated hypercalcemia normalized their serum calcium values after infusions of zoledronic acid (4 mg) and pamidronate (90 mg), respectively. Zoledronic acid produced a longer duration of response when treating hypercalcemia of malignancy—32 days versus 18 days for pamidronate. The nadir of serum calcium does not occur until 4 to 5 days after administration of either agent. Retreatment with either agent can be conducted after recurrence of hypercalcemia. Transient fever and myalgia occur in between 10% and 20% of patients who undergo intravenous bisphosphonate therapy. Increases in serum creatinine of greater than or equal to 0.5 mg/dL occur in about 15% of patients. Intravenous bisphosphonates should be used cautiously (if at all) and at reduced doses when the baseline serum creatinine exceeds 2.5 mg/dL.

In patients with severe hypercalcemia and in those with renal insufficiency that is refractory to rehydration, it may be necessary to use an alternative second-line antiresorptive agent for a few days while awaiting the full therapeutic effect of intravenous bisphosphonates. For this purpose, synthetic salmon CT may be administered at a dose of 4 to 8 IU/kg subcutaneously every 12 hours. This is a useful adjunct acutely, but most patients become refractory to CT within days to weeks, so it is not suitable for chronic use.

In 33 patients with malignancy-associated hypercalcemia refractory to intravenous bisphosphonates (serum $[Ca^{2+}]$ >12.5 mg/dL), the neutralizing monoclonal antibody to receptor activator of nuclear factor kappa B ligand (RANK-L) denosumab demonstrated efficacy at controlling hypercalcemia. Patients were given denosumab (120 mg subcutaneously) on days 1, 8, 15, and 29 and then every 4 weeks. Using this treatment protocol, by day 10, 64% of patients achieved an albumin-corrected serum $[Ca^{2+}]$ of less than or equal to 11.5 mg/dL, while 33% of patients declined to less than or equal to 10.8 mg/dL. Throughout the study, 70% and 64% of patients achieved corrected serum $[Ca^{2+}]$ values of less than or equal to 11.5 mg/dL and less than or equal to 10.8 mg/dL, respectively, with a mean duration of response estimated to be 104 days. Thus, this agent is an additional therapy for refractory hypercalcemia in such patients.

Glucocorticoid administration is first-line treatment for hypercalcemia in patients with multiple myeloma, lymphoma, sarcoidosis, or intoxication with vitamin D or vitamin A. Glucocorticoids are also beneficial in some patients with breast carcinoma. However, they are of little use in most other patients with solid tumors and hypercalcemia.

HYPOCALCEMIA

Classification

Both PTH and $1,25(OH)_2D$ maintain a normal serum calcium and are thus central to the defense against hypocalcemia. Hypocalcemic disorders are best understood as failures of the adaptive response. Thus, chronic hypocalcemia can result from a failure to secrete PTH, altered responsiveness to PTH, a deficiency of vitamin D, or a resistance to vitamin D (Table 8–7). Acute hypocalcemia is most often the consequence of an overwhelming challenge to the adaptive response such as rhabdomyolysis, in which a flood of phosphate from injured skeletal muscle inundates the extracellular fluid.

Clinical Features

Most of the symptoms and signs of hypocalcemia occur because of increased neuromuscular excitability (tetany, paresthesias, muscle cramping, seizures, organic brain syndrome), and many of the complications of treatment of hypocalcemia result from the deposition of calcium in soft tissues (cataract, calcification of basal ganglia, nephrolithiasis, nephrocalcinosis, reduced renal function).

A. Neuromuscular manifestations Clinically, the hallmark of severe hypocalcemia is tetany. Tetany is a state of spontaneous tonic muscular contraction. Overt tetany is often heralded by tingling paresthesias in the fingers and around the mouth, but the classic muscular component of tetany is carpopedal spasm. This begins with adduction of the thumb, followed by flexion of the metacarpophalangeal joints, extension of the interphalangeal joints, and flexion of the wrists to produce the *main d'accoucheur* posture (Figure 8–18). These involuntary muscle contractions are painful. Although the hands are most typically involved, tetany can involve other muscle groups, including life-threatening spasm of laryngeal muscles (laryngospasm). Electromyographically, tetany is typified by repetitive motor neuron action potentials, usually grouped as doublets. Tetany is not specific for hypocalcemia. It also occurs with

TABLE 8–7   Causes of hypocalcemia.

Hypoparathyroidism
Surgical
Idiopathic
Neonatal
Familial
Autoimmune
Deposition of metals (iron, copper, aluminum)
Postradiation
Infiltrative
Functional (in hypomagnesemia)

Resistance to PTH Action
Pseudohypoparathyroidism
Renal insufficiency
Medications that block osteoclastic bone resorption
   Plicamycin
   Calcitonin
   Bisphosphonates
   Denosumab

Failure to Produce 1,25(OH)$_2$D Normally
Vitamin D deficiency
Hypoparathyroidism
Diseases associated with increased FGF23 levels
Pseudovitamin D deficiency rickets (inactivating *CYP27B1* mutations)

Resistance to 1,25(OH)$_2$D Action
Hereditary vitamin D-resistant rickets (inactivating *VDR* mutations)

Acute Complexation or Deposition of Calcium
Acute hyperphosphatemia
   Crush injury with myonecrosis
   Rapid tumor lysis
   Parenteral phosphate administration
   Excessive enteral phosphate
      Oral (phosphate-containing antacids)
      Phosphate-containing enemas
Acute pancreatitis
Citrated blood transfusion
Rapid, excessive skeletal mineralization
   Hungry bone syndrome
   Osteoblastic metastasis
   Vitamin D therapy for vitamin D deficiency

hypomagnesemia and metabolic alkalosis, and the most common cause of tetany is respiratory alkalosis from hyperventilation.

Lesser degrees of neuromuscular excitability (eg, serum [Ca] 7.5-8.5 mg/dL) produce latent tetany, which can be elicited by
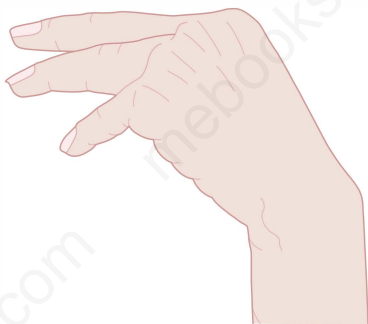
FIGURE 8–18   Position of fingers in carpal spasm due to hypocalcemic tetany. (Reproduced with permission from Ganong WF. *Review of Medical Physiology.* 16th ed. New York: McGraw-Hill Education; 1993.)

testing for Chvostek and Trousseau signs. Chvostek sign is elicited by tapping the facial nerve about 2 cm anterior to the earlobe, just below the zygoma. The response is a contraction of facial muscles ranging from twitching of the angle of the mouth to hemifacial contractions. The specificity of the test is low; about 25% of normal individuals have a mild Chvostek sign. Trousseau sign is elicited by inflating a blood pressure cuff to about 20 mm Hg above systolic pressure for 3 minutes. A positive response is carpal spasm. Trousseau sign is more specific than Chvostek, but 1% to 4% of normal individuals have positive Trousseau signs.

Hypocalcemia predisposes to focal or generalized seizures. Other central nervous system effects of hypocalcemia include pseudotumor cerebri, papilledema, confusion, lassitude, depression, organic brain syndrome, and coma. Twenty percent of children with chronic hypocalcemia develop mental retardation. The basal ganglia are often calcified in patients with longstanding hypoparathyroidism or PHP. This is usually asymptomatic but may produce a variety of movement disorders.

B. Other manifestations of hypocalcemia

1. **Cardiac effects**—Repolarization is delayed, with prolongation of the QT interval. Excitation-contraction coupling may be impaired, and refractory congestive heart failure is sometimes observed, particularly in patients with underlying cardiac disease.
2. **Ophthalmologic effects**—Subcapsular cataract is common in chronic hypocalcemia, and its severity is correlated with the duration and level of hypocalcemia.
3. **Dermatologic effects**—The skin is often dry and flaky and the nails brittle.

CAUSES OF HYPOCALCEMIA

1. HYPOPARATHYROIDISM

Hypoparathyroidism may be surgical, autoimmune, familial/genetic, or idiopathic. The signs and symptoms are those of chronic hypocalcemia. Biochemically, the hallmarks of hypoparathyroidism are hypocalcemia, hyperphosphatemia (because the phosphaturic effect of PTH is lost), and an inappropriately low or undetectable PTH level.

Surgical Hypoparathyroidism

The most common cause of hypoparathyroidism is surgery on the neck, with removal or destruction of the parathyroid glands. The operations most often associated with hypoparathyroidism are cancer surgery (typically for laryngeal cancer), total thyroidectomy, and parathyroidectomy, but the skill and experience of the surgeon are more important predictors than the nature of the operation. Tetany ensues 1 or 2 days postoperatively or even earlier, and about half of patients with postoperative tetany recover sufficiently so they do not require long-term replacement therapy. In these cases, a devitalized parathyroid remnant has recovered its blood supply and resumes secretion of PTH. In some patients,

hypocalcemia may not become evident until years after the procedure.

In patients with severe hyperparathyroid bone disease preoperatively, a syndrome of postoperative hypocalcemia can follow successful parathyroidectomy. This is the "**hungry bone syndrome**," which results from avid uptake of calcium and phosphate by the bones. The parathyroids, although intact, cannot compensate. The syndrome is usually seen in patients with an elevated preoperative serum alkaline phosphatase and/or severe uremic secondary hyperparathyroidism. It can usually be distinguished from surgical hypoparathyroidism by the serum phosphorus, which is low in the hungry bones syndrome, because of skeletal avidity for phosphate, and high in hypoparathyroidism, and by the serum PTH, which becomes appropriately elevated in the hungry bone syndrome. This syndrome can be seen under any condition in which high levels of bone resorption are treated with an antiresorptive agent or even after thyroidectomy for severe hyperthyroidism.

Idiopathic Hypoparathyroidism

Acquired hypoparathyroidism is sometimes seen in the setting of polyendocrinopathies. Most commonly, it is associated with primary adrenal insufficiency and mucocutaneous candidiasis in the syndrome of *autoimmune polyendocrine syndrome type 1 (APS1)* (see Chapter 3). The typical age at onset of hypoparathyroidism is 5 to 9 years. A similar form of hypoparathyroidism can occur as an isolated finding. The age at onset of idiopathic hypoparathyroidism is 2 to 10 years, and there is a preponderance of female cases. Circulating autoantibodies are common in both APS1 and in isolated hypoparathyroidism. Many patients have antibodies that recognize the CaSR or react with NALP5 (NACHT leucine-rich-repeat protein 5), an intracellular signaling molecule expressed in the parathyroid. More important from the standpoint of pathogenesis are autoantibodies to type 1 interferons (IFNs, alpha and omega). Antibodies to these IFNs are present in nearly 100% of patients with APS1, often years before disease presentation, and can persist for years, making them the best markers for the disorder. Antibodies to interleukin (IL)-17 and IL-22 are also highly prevalent in APS1 patients. When tested in lab-based assays, these anticytokine antibodies (anti-IFN, anti-IL-17, IL-22) have neutralizing properties against these cytokines. However, exactly how these antibodies contribute to the pathogenesis of the immunologic alterations in APS1 patients *in vivo* is not completely clear. APS1 is an autosomal recessive (or rarely dominant) disorder due to mutations in the *AIRE* (autoimmune regulator) gene. AIRE is involved in the complex process of immune tolerance to self and is a key transcription factor in medullary thymic epithelial cells. AIRE promotes expression of thousands of tissue antigens enabling T cells autoreactive to these antigens to be deleted in the process of negative selection.

Another form of autoimmune hypoparathyroidism has been reported in patients with autoantibodies to the CaSR that functionally activate it and suppress PTH secretion. These rare cases have been detected along with other autoimmune disorders such as Addison disease and Hashimoto thyroiditis.

Familial Hypoparathyroidism

Hypoparathyroidism is an uncommon disorder and presents in familial forms transmitted as either autosomal dominant or recessive or X-linked traits. Autosomal recessive hypoparathyroidism occurs in families with *PTH* gene mutations that interfere with the normal processing of PTH. Another autosomal recessive form of this disease is due to a deletion of the 5' sequence of the gene encoding the transcription factor *glial cell missing 2*, which is necessary for parathyroid gland development. Hypoparathyroidism is evident in the newborn period because of parathyroid gland agenesis.

Autosomal dominant hypoparathyroidism (ADH) type 1 is due to point mutations in the *CaSR* gene, which render the protein constitutively active. This property enables the receptor to mediate suppression of PTH secretion at normal and even subnormal serum calcium levels. Many mutations have been described; affected individuals typically have mild hypoparathyroidism (mildly reduced serum $[Ca^{2+}]$ and PTH levels). These patients also demonstrate marked hypercalciuria due to mutant CaSRs in their kidneys. The set-point for calcium-induced suppression of PTH secretion in these patients is shifted to the left, as is the effect of serum $[Ca^{2+}]$ on renal calcium excretion. ADH type 2 is due to gain-of-function mutations in gene encoding G-alpha 11, which couples to the CaSR leading to enhanced suppression of PTH secretion at even low serum calcium levels. Thus, ADH types 1 and 2 are the mirror images of FHH types 1 and 2. Therapy is often not necessary and risks precipitating even greater degrees of hypercalciuria, nephrocalcinosis, and renal stones.

Other Causes of Hypoparathyroidism

Neonatal hypoparathyroidism can be part of the **DiGeorge syndrome** (dysmorphic facies, cardiac defects, immune deficiency, and hypoparathyroidism) due to a microdeletion (or other genetic abnormalities) on chromosome 22q11.2; the **HDR syndrome** (hypoparathyroidism, sensorineural deafness, and renal anomalies) due to mutations in or deletion of an allele of the *GATA3* transcription factor; and other rare conditions. Transfusion-dependent individuals with thalassemia or red cell aplasia who survive into the third decade of life are susceptible to hypoparathyroidism as the result of iron deposition in the parathyroid glands. Copper deposition can cause hypoparathyroidism in Wilson disease. Infiltration with metastatic carcinoma is a very rare cause of hypoparathyroidism.

Severe magnesium depletion temporarily paralyzes the parathyroid glands, preventing secretion of PTH. Magnesium depletion also blunts the actions of PTH on target organs (kidney and bone) to counteract the hypocalcemia. This is seen with magnesium losses due to gastrointestinal and renal disorders and chronic alcoholism. The syndrome responds immediately to infusion of magnesium. As discussed earlier in the section on regulation of PTH secretion, magnesium is required for effective stimulus-secretion coupling in the parathyroids. Several inherited disorders of renal tubular dysfunction lead to the loss of magnesium and can cause hypomagnesemia and hypocalcemia.

2. PSEUDOHYPOPARATHYROIDISM

PHP is a genetic disorder of target-organ unresponsiveness to PTH. Biochemically, it mimics hormone-deficient forms of hypoparathyroidism, with hypocalcemia and hyperphosphatemia, but the PTH level is elevated, and there is a markedly blunted response to the administration of PTH (see Diagnosis discussed later).

Clinical Features

Two distinct forms of PHP are recognized. PHP type IB is a disorder of isolated resistance to PTH, which presents with the biochemical features of hypocalcemia, hyperphosphatemia, and secondary hyperparathyroidism. PHP type IA has, in addition to these biochemical features, a characteristic somatic phenotype known as **Albright hereditary osteodystrophy (AHO)**. This consists of short stature, round face, short neck, obesity, brachydactyly (short digits), shortened metatarsals, subcutaneous ossifications, and often reduced intelligence. Because of shortening of the metacarpal bones—most often the fourth and fifth metacarpals—affected digits have a dimple, instead of a knuckle, when a fist is made (Figure 8–19). Primary hypothyroidism is frequently seen. Less commonly, these patients have abnormalities of reproductive function—oligomenorrhea in females and infertility in males due to primary hypogonadism. Interestingly, certain individuals in families with PHP inherit the somatic phenotype of AHO without any disorder of calcium metabolism; this state, which mimics PHP, is called **pseudo-PHP** or **PPHP**.

Pathophysiology

PHP type IA is caused by the loss of function of one allele (haploinsufficiency) of the gene encoding the stimulatory G protein alpha subunit, (G_s *alpha* or *GNAS*). This abnormality is predicted to produce only 50% of the normal levels of the alpha subunit of the heterotrimeric G_s, which couples the PTH receptor to adenylyl cyclase. Patients with PHP type IA have a markedly blunted response of urinary cAMP to administration of PTH. Because G_s also couples many other receptors to adenylyl cyclase, the expected result of this mutation would be a generalized disorder of hormonal unresponsiveness. The presence of primary hypothyroidism and primary hypogonadism in these patients indicates that resistance to TSH, LH, and FSH occurs fairly commonly. Responsiveness to other hormones (eg, ACTH, glucagon) is fairly normal. Thus, a 50% loss of the G_s alpha protein produces resistance to some hormones, but not others. *GNAS* is also mutated in individuals with PPHP, who have AHO but normal responsiveness to PTH. Thus, a mutation in *GNAS* invariably produces AHO but only sometimes produces resistance to PTH, suggesting that the occurrence of a metabolic phenotype is determined by other factors described later (Table 8–8).

In PHP type IB, there is resistance to PTH but no somatic phenotype (ie, no AHO). The disorder is linked to the *GNAS* locus but is not due to mutations in the coding region of *GNAS*. Loss of parent-specific imprinting patterns underlie PHP IB. Epigenetic defects in *GNAS* cause the differentially methylated region at exon A/B to lose its imprinting function. Most cases demonstrate a maternally derived 3-kb deletion of DNA sequences

FIGURE 8–19   Hands of a patient with pseudohypoparathyroidism. **A.** Note the short fourth fingers. **B.** Note the "absent" fourth knuckle. **C.** Film shows the short fourth metacarpal. (Reproduced with permission from Stanbury JB, Wyngaarden JB, Fredrickson DS. *The Metabolic Basis of Inherited Disease*. 4th ed. New York: McGraw-Hill Education; 1978.)

TABLE 8–8 Features of GNAS spectrum disorders.

	PHP IA	PPHP[a]	PHP IB	POH[b]
Hypocalcemia	Yes	No	Yes	No
Response to PTH	No	Yes	No	Yes
Albright hereditary osteodystrophy	Yes	Yes	No	No[c]
GNAS coding region mutation	Yes	Yes	No	Yes
GNAS regulatory region mutation or deletion	No	No	Yes	No
Generalized target organ unresponsiveness	Yes	No	No	No

[a]Pseudopseudohypoparathyroidism.
[b]POH, progressive osseous heteroplasia.
[c]heterotopic ossifications in soft tissues only.

located more than 200 kb centromeric of the *GNAS* locus at the *STX16* locus. Other deletions remove this entire differentially methylated region—a key regulator of the levels of *GNAS* transcription. Transcription directed by this (unmethylated) promoter does not produce appropriate levels of G$_s$ alpha protein in the renal cortex where this biallelic promoter activity is important. Thus, abnormal methylation of *GNAS* regulatory sequences plays a key role in the pathogenesis of PHP IB.

Progressive osseous heteroplasia (POH) is another rare disorder that involves the *GNAS* locus (see Table 8–8). Ectopic bone formation occurs in this disorder, similar but to a more severe extent than in PHP IA. POH begins in early childhood with ectopic bone forming in the dermis, muscles, and connective tissues. Features can include AHO (short stature, brachydactyly), but there are no calcium or PTH abnormalities. POH is considered part of the spectrum of *GNAS* disorders.

Genetics

PHP type IA is inherited as an autosomal dominant trait. Individuals who have acquired the trait from their fathers almost always present with PPHP and lack hormone resistance. When the mutant allele is derived from the mother, PHP with hormone resistance is always present. This pattern of inheritance is due to genomic imprinting of the *GNAS* locus. In the kidney cortex, the maternal allele is preferentially expressed, and the paternal allele is silenced. Thus, if the mutant *GNAS* allele is maternally derived, the resulting offspring will have PHP; if the mutant *GNAS* allele is paternally derived (and therefore silenced), the resulting offspring will have PPHP. Tissue-specific imprinting of the *GNAS* gene, therefore, determines the expression of hormone resistance in the kidney and thyroid. The expression of AHO does not depend on imprinting mechanisms and occurs in PHP type IA and PPHP.

Most cases of PHP IB are sporadic. The few inherited cases that haven characterized are the result of deletions of maternally-derived cis-acting control elements in the *GNAS* locus. Additionally, there are cases of uniparental isodisomy described. POH, in contrast to PHP, is due to paternal loss of function mutations in *GNAS*.

Diagnosis

Several disorders present with hypocalcemia and secondary hyperparathyroidism (eg, vitamin D deficiency), but when these features occur together with hyperphosphatemia or AHO, this suggests the diagnosis of PHP. To confirm that resistance to PTH is present, the patient can be challenged with PTH (the Ellsworth-Howard test). For this purpose, synthetic human PTH(1-34) (teriparatide acetate, 3 IU/kg body weight) is infused intravenously over 10 minutes during a water diuresis, and urine is collected during the hour preceding the infusion, during the half-hour following the infusion, 30 to 60 minutes after the infusion, and 1 to 2 hours after the infusion and assayed for cAMP and creatinine. Data are expressed as nanomoles of cAMP per liter of glomerular filtrate, based on creatinine measurements. Normally, there is an increase in urinary cAMP of more than 300 nmol/L glomerular filtrate after administration of PTH. The use of the urinary phosphate response as a gauge of PTH responsiveness is much less reliable. For practical purposes, hypocalcemia, hyperphosphatemia, elevated PTH levels, and normal levels of vitamin D metabolites (with or without AHO) confirm the presence of resistance to PTH.

3. VITAMIN D DEFICIENCY

Pathogenesis

Vitamin D deficiency results from one or more factors: inadequate sunlight exposure, inadequate nutrition, or malabsorption. In addition, drugs that activate the catabolism of vitamin D and its metabolites, such as phenytoin and phenobarbital, can precipitate vitamin D deficiency in subjects with marginal vitamin D status. Although the human skin is capable of producing sufficient amounts of vitamin D if exposed to sunlight of adequate intensity, institutionalized and home-bound patients frequently do not get adequate exposure. Furthermore, the fear of skin cancer has led many to avoid sunlight exposure or to apply protective agents that block the ultraviolet portion of sunlight from reaching the lower reaches of the epidermis where most of the vitamin D is produced. Heavily pigmented and elderly individuals have less efficient production of vitamin D for a given exposure to ultraviolet irradiation. Serum levels of 25(OH)D are used to indicate vitamin D status. Although there is no consensus on what level of 25(OH)D constitutes sufficiency, 30 ng/mL has become the preferred target in North America by many experts. Recent recommendations from an expert panel formed by the Institute of Medicine, however, concluded that 20 ng/mL was "sufficient." Using the level of 30 ng/mL as a standard, a large portion of the population is vitamin D-insufficient. A recent NHANES survey found that in the population over age 60, 67% of Caucasians, 82% of Hispanics, and 88% of African Americans had 25(OH)D levels below 30 ng/mL. Studies of hospitalized patients or those in nursing homes show almost universal vitamin D insufficiency, if not frank deficiency. The intensity of sunlight is an important factor that limits effective vitamin D production as a function of season (summer > winter) and latitude (less intense the higher the latitude). The supplementation of milk has reduced the incidence of vitamin D

deficiency in the United States, but most countries do not follow this practice. Even in the United States, severe vitamin D deficiency and rickets may occur in children of vegetarian mothers who avoid milk products (and presumably have reduced vitamin D stores) and in children who are not weaned to vitamin D-supplemented milk by age 2. Breast milk contains little vitamin D, although vitamin D content can be increased by supplementing the mother. Adults who avoid milk as well as sunlight are likewise at risk. Individuals with a variety of small bowel, pancreatic, and biliary tract disease and those after partial gastrectomy or intestinal bypass surgery have reduced capacity to absorb the vitamin D from the diet.

Clinical Features

The clinical features of individuals with severe vitamin D deficiency will be discussed more thoroughly in the section on "Osteomalacia and Rickets." Severe vitamin D deficiency should be suspected in individuals complaining of lethargy, proximal muscle weakness, and bone pain who have frankly low or low normal serum calcium and phosphate and low urine calcium on routine biochemical evaluation. A low serum 25(OH)D level is diagnostic in this setting. $1,25(OH)_2D$ levels are often normal and reflect the increased 1-hydroxylase activity in these subjects that is responding appropriately to the increased PTH levels as well as the low serum calcium and phosphate levels. However, there is increasing awareness that less severe degrees of vitamin D deficiency may not present with obvious musculoskeletal signs and symptoms. These individuals may be more susceptible to conditions such as hyperparathyroidism, osteoporosis, increased risk of falls and fractures, but also increased infections, hypertension, increased cardiovascular disease, diabetes mellitus, and various malignancies. Convincing data from large-scale randomized placebo controlled clinical trials demonstrating cause and effect for vitamin D deficiency and many of these disorders are lacking, although animal studies and a large body of epidemiologic evidence and case control studies point in this direction.

Treatment

The goal in treating severe vitamin D deficiency manifesting as rickets or osteomalacia is to normalize the clinical, biochemical, and radiologic abnormalities without producing hypercalcemia, hyperphosphatemia, hypercalciuria, nephrolithiasis, or ectopic calcification. To realize this goal, patients must be followed carefully. As the bone lesions heal or the underlying disease improves, the dosage of vitamin D, calcium, or phosphate needs to be adjusted to avoid such complications. With the appreciation that most individuals have a less severe form of vitamin D deficiency and do not have obvious signs or symptoms associated with rickets or osteomalacia, the goal of treatment is to achieve a level of 25(OH)D that existing data indicate will reduce the predisposition of such individuals to the numerous diseases associated with vitamin D deficiency. Correction of vitamin D deficiency can usually be achieved by the weekly administration of 50,000 IU ergocalciferol for 6 to 8 weeks followed by replacement doses of 800 to 2000 IU/d based on maintenance of adequate serum 25(OH)D levels. Patients with malabsorption including patients

after bariatric surgery may require larger amounts of vitamin D. 25(OH)D (50-100 μg/d) is better absorbed than vitamin D and may be used, if malabsorption of vitamin D is a limiting factor, but this metabolite of vitamin D is not readily available in the United States. Calcitriol $[1,25(OH)_2D]$ and its analogs are not appropriate therapy for patients with vitamin D deficiency because of the likely requirement for vitamin D metabolites other than $1,25(OH)_2D$ and the advantage of providing adequate substrate [25(OH)D] for tissues capable of producing their own $1,25(OH)_2D$ as the need arises (eg, as in the immune response). The exception to this rule is in patients with renal failure who are incapable of producing adequate $1,25(OH)_2D$. Vitamin D therapy should be combined with calcium based on current recommendations for daily calcium intake (see the section "Nonpharmacologic Aspects of Osteoporosis Management" and Table 8–12). Care must be taken in managing patients with vitamin D deficiency who also have elevated PTH levels, because longstanding vitamin D deficiency may produce a degree of autonomy in the parathyroid glands such that rapid replacement with calcium and vitamin D could result in hypercalcemia or hypercalciuria. A listing of available vitamin D metabolites and analogs with their main indications for clinical use is presented in Table 8–2.

4. PSEUDOVITAMIN D DEFICIENCY

Pseudovitamin D deficiency is a rare autosomal recessive disease in which rickets is accompanied by low levels of $1,25(OH)_2D$ but normal levels of 25(OH)D. The disease is due to mutations in the *25(OH)D 1-hydroxylase* gene (*CYP27B1*) that render it nonfunctional. Both alleles need to be defective in order for the disease to be manifest. Although affected patients do not respond to doses of vitamin D that are effective in subjects with vitamin D deficiency, they can respond to pharmacologic doses of vitamin D and to physiologic doses of calcitriol, which is the preferred treatment.

5. HEREDITARY VITAMIN D–RESISTANT RICKETS

Hereditary vitamin D–resistant rickets is a rare autosomal recessive disease that presents in childhood with rickets similar to that seen in patients with vitamin D deficiency. Many of these patients also have alopecia, which is not characteristic of vitamin D deficiency or pseudovitamin D deficiency. The biochemical changes are similar to those reported in subjects with vitamin D deficiency except that the $1,25(OH)_2D$ levels are generally very high. The disease is caused by inactivating mutations in the *VDR*. The location of the mutation can affect the severity of the disease. In particular, mutations in the DNA-binding domain (exons 2, 3) that prevent DNA binding and mutations in the ligand-binding domain (exons 5-7) that prevent $1,25(OH)_2D$ binding lead to a more severe phenotype. The latter mutations may alter but not inhibit DNA/ligand binding. These patients are treated with large doses of calcitriol and dietary calcium and may show partial or complete remission as they grow older. An animal

model of this disease (inactivation or knockout of both *VDR* alleles) demonstrates that the bone disease can be corrected with high dietary intake of calcium and phosphate, although the alopecia is not altered. This disease points to a role for the VDR in epidermis and hair development that is independent of its activity in the intestine and bone.

6. OTHER HYPOCALCEMIC DISORDERS

Hypoalbuminemia produces a low total serum $[Ca^{2+}]$ because of a reduction in the bound fraction of calcium, but the ionized $[Ca^{2+}]$ is normal. The ionized $[Ca^{2+}]$ can be determined directly, or the effect of hypoalbuminemia can be roughly corrected for by using the following formula:

$$Corrected\ serum\ calcium = Measured\ serum\ calcium + (0.8)$$
$$(4 - Measured\ serum\ albumin)$$

Thus, in a patient with a serum $[Ca^{2+}]$ of 7.8 mg/dL and a serum albumin of 2 g/dL, the corrected serum $[Ca^{2+}]$ is 7.8 + (0.8)(4 − 2) = 9.4 mg/dL.

Several disorders produce acute hypocalcemia even if homeostatic mechanisms are intact, simply because they overwhelm these mechanisms. Acute hyperphosphatemia resulting from rhabdomyolysis or tumor lysis, often in the setting of renal insufficiency, may produce severe symptomatic hypocalcemia. Transfusion of large volumes of citrated blood causes acute hypocalcemia by complexation of calcium as calcium citrate. In this instance, total calcium may be normal, but the ionized fraction is reduced. In acute pancreatitis, hypocalcemia is an ominous prognostic sign. The mechanism of hypocalcemia is sequestration of calcium by saponification with fatty acids, which are produced in the retroperitoneum by the action of pancreatic lipases. Skeletal mineralization, when very rapid, can cause hypocalcemia. This is seen in the **hungry bone syndrome**, which was discussed earlier in the section on surgical hypoparathyroidism, and occasionally with widespread osteoblastic metastases from prostatic carcinoma.

TREATMENT OF HYPOCALCEMIA

Acute Hypocalcemia

Patients with tetany should receive intravenous calcium preferably as calcium gluconate (90 mg calcium/10 mL) or calcium gluceptate (90 mg calcium/10 mL) over 10 to 20 minutes. Approximately 200 mg of elemental calcium can be given over 20 to 40 minutes by slow intravenous infusion with constant electrocardiographic monitoring. The patient must be observed for stridor and the airway secured if necessary. Oral calcium and a rapidly acting preparation of vitamin D (eg, calcitriol) should be started. If necessary, calcium can be infused in doses of 400 to 1000 mg/24 h until oral therapy has taken effect. Intravenous calcium is irritating to the veins and is best infused into a large vein or through a central venous catheter.

Chronic Hypocalcemia

The objective of chronic therapy is to keep the patient free of symptoms and to maintain a serum $[Ca^{2+}]$ of approximately 8.5 to 9.0 mg/dL. With lower serum $[Ca^{2+}]$, the patient may not only experience symptoms but may be predisposed over time to cataract formation if the phosphate level is also high. With serum calcium concentrations in the upper normal range, there may be marked hypercalciuria, which occurs because the hypocalciuric effect of PTH has been lost. This may predispose to nephrolithiasis, nephrocalcinosis, and chronic renal disease. In addition, the patient with borderline elevated calcium is at increased risk of overshooting the therapeutic goal and may develop symptomatic hypercalcemia.

The mainstays of treatment are calcium and a form of vitamin D. Oral calcium can be given in a dose of 1.5 to 3 g of elemental calcium or more per day. These large doses of calcium reduce the doses of vitamin D that are needed and allow for rapid normalization of serum calcium if vitamin D intoxication subsequently occurs. Numerous preparations of calcium are available. A short-acting preparation of vitamin D (calcitriol) and the very long-acting preparations such as vitamin D_2 (ergocalciferol) are available and can be used (see Table 8–2). Although regimens that use ergocalciferol are economical, they have the disadvantage of the long half-life of vitamin D_2 if vitamin D intoxication occurs. Additionally, high doses of vitamin D_2 are needed to treat hypoparathyroidism, if that is the agent is to be used. In PTH deficiency states, the effects of endogenous PTH on renal 1-alpha hydroxylase activity are missing, and there is little conversion of $25(OH)D$ to $1,25(OH)_2D$. Caution must be exercised in the introduction of other drugs that influence calcium metabolism. For example, thiazide diuretics have a hypocalciuric effect. By reducing urinary calcium excretion in treated patients, whose other adaptive mechanisms such as the modulation of PTH secretion by calcium, are nonoperative and who are thus absolutely dependent on renal excretion of calcium to maintain the serum calcium level, thiazides may produce significant hypercalcemia. In a similar way, intercurrent illnesses that compromise renal function (and therefore calcium excretion) may produce dangerous hypercalcemia in the patient who is maintained on large doses of vitamin D. Short-acting preparations of vitamin D (eg, calcitriol or $1,25(OH)_2D$ and 1-alpha calcidiol) are less prone to some of these effects but may require more frequent titration and are more expensive than vitamin D_2.

BONE ANATOMY AND REMODELING

FUNCTIONS OF BONE

Bone has four major functions:

1. It provides rigid support to extremities and body cavities containing vital organs. In diseases in which the skeleton is weakened or defective, erect posture may be impossible and vital

organ function may be compromised. An example is the cardiopulmonary dysfunction that occurs in patients with severe kyphosis due to vertebral collapse.

2. Bones are crucial to locomotion in that they provide efficient levers and sites of attachment for muscles. With bony deformity, these levers become defective, and severe abnormalities of gait develop.

3. Bone provides a large reservoir of ions, such as calcium, phosphate, magnesium, and sodium that are critical for life and can be mobilized when the external environment fails to provide them.

4. Bone houses the hematopoietic elements. There is increasing evidence of a trophic relationship between the stromal cells in bone and the hematopoietic elements.

STRUCTURE OF BONE

The material properties of normal bone strike an ideal balance between rigidity and elasticity. Bone is rigid enough to provide structural stability and resist applied forces but not overly mineralized and brittle, which would result in an increased tendency to fracture. Bone must also be light enough to be moved by muscle contractions. Cortical bone, composed of densely packed layers of mineralized collagen, provides rigidity and is the major component of tubular bones (Figure 8–20). Trabecular (cancellous) bone is spongy in cross-section, provides strength and elasticity, and constitutes the major portion of the axial skeleton. Disorders in which cortical bone is defective or reduced in mass lead to
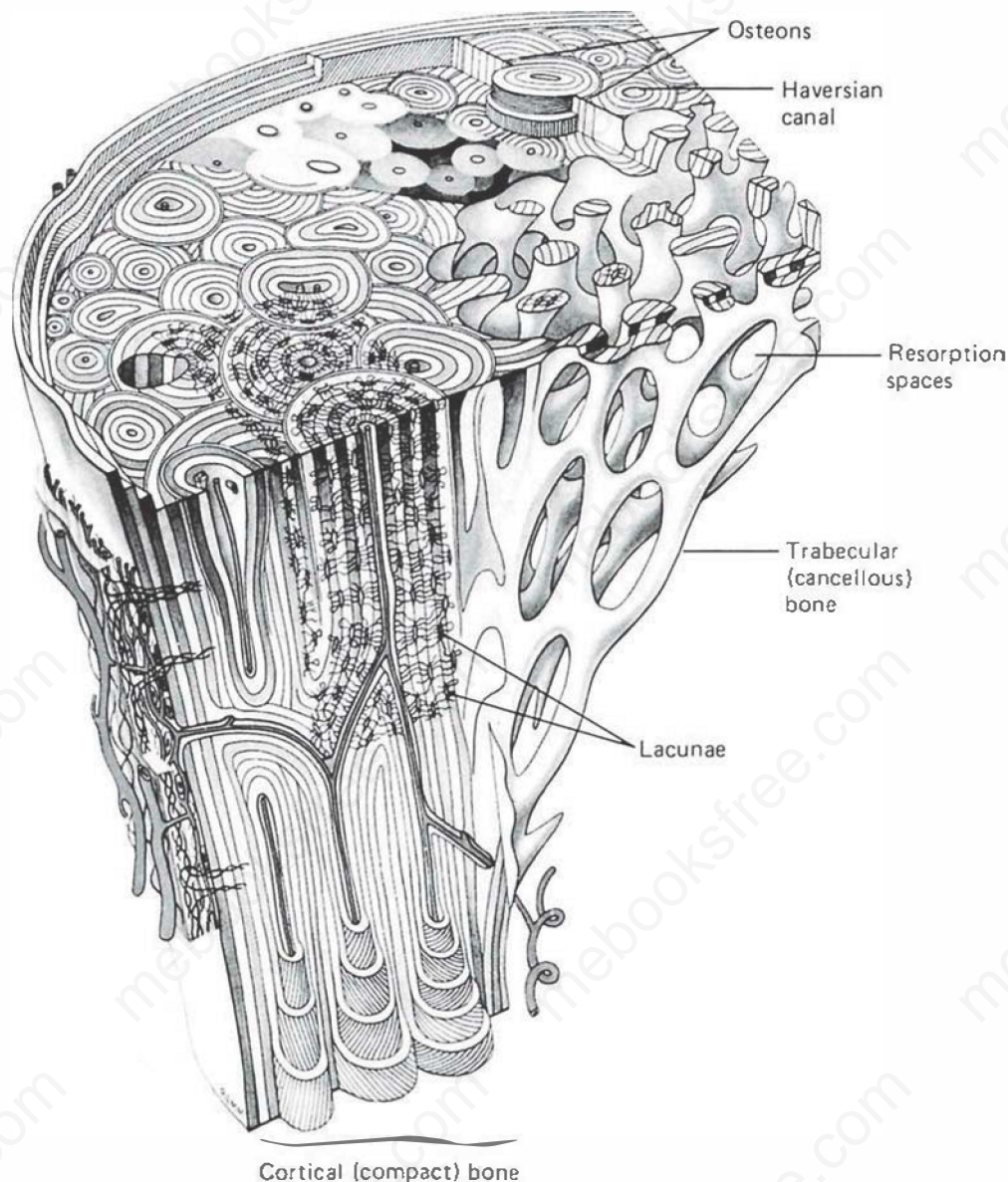
FIGURE 8–20 Diagram showing features of the microstructure of mature bone seen in both transverse (top) and longitudinal section. Areas of cortical (compact) and trabecular (cancellous) bone are included. (Reproduced with permission from Standring S. *Gray's Anatomy*. 40th ed. Churchill Livingstone/Elsevier; 2008.)

fractures of the long bones. Disorders in which trabecular bone is defective or scanty, in contrast, lead preferentially to vertebral fractures. Fractures of long bones may also occur because normal trabecular bone reinforcement is lost.

Two-thirds of the weight of bone is due to mineral; the remainder is due to water and type I collagen. Minor organic components such as proteoglycans, lipids, acidic proteins containing gamma-carboxyglutamic acid, osteonectin, osteopontin, and growth factors comprise the remainder.

Bone Mineral

The mineral of bone is present in two forms. The major form consists of hydroxyapatite in crystals of varying maturity. The remainder is amorphous calcium phosphate, which lacks a coherent x-ray diffraction pattern, has a lower calcium-to-phosphate ratio than pure hydroxyapatite, occurs in regions of active bone formation, and is present in larger quantities in young bone.

Bone Cells

Bone is composed of three types of cells: the osteoblast, the osteocyte, and the osteoclast.

A. Osteoblast   The osteoblast is the principal bone-forming cell. It arises from a pool of mesenchymal stem cells in the bone marrow, which as they differentiate, acquire a set of characteristics, including expression of PTH and vitamin D receptors; surface expression of the ectoenzyme alkaline phosphatase; and expression of bone matrix protein genes—type I collagen, osteocalcin, osteopontin, and others. Differentiated osteoblasts are directed to the bone surface, where they line regions of new bone formation, laying down bone matrix (osteoid) in orderly lamellae and inducing its mineralization (Figure 8–21). In the

FIGURE 8–21   The remodeling cycle. **A.** Resting trabecular surface. **B.** Multinucleated osteoclasts dig a cavity of approximately 20 microns. **C.** Completion of resorption to 60 microns by mononuclear phagocytes. **D.** Recruitment of osteoblast precursors to the base of the resorption cavity. **E.** Secretion of new matrix (gray shading) by osteoblasts. **F.** Continued secretion of matrix, with initiation of calcification (black areas). **G.** Completion of mineralization of new matrix. Bone has returned to quiescent state, but a small deficit in bone mass persists.

mineralization process, hydroxyapatite crystals are deposited on the collagen layers to produce lamellar bone. Mineralization requires an adequate supply of extracellular calcium and phosphate as well as alkaline phosphatase, which is secreted in large amounts by active osteoblasts. The fate of senescent osteoblasts is not well defined. Some probably become flattened, inactive lining cells on trabecular bone surfaces, some are buried in cortical bone as osteocytes, and others undergo apoptosis.

B. Osteocyte   Osteoblasts that remain within cortical bone during the remodeling process become osteocytes. Protein synthetic activity decreases markedly, and the cells develop multiple processes (canaliculi) that reach out through lacunae in bone tissue to communicate with nutrient capillaries, with processes of other osteocytes within a unit of bone (osteon), and also with the cell bodies of surface osteoblasts (see Figure 8–20). The physiologic functions of osteocytes are varied. They are believed to (1) act as a cellular syncytium that permits the translocation of mineral in and out of regions of bone removed from surfaces; (2) serve as the sensors of mechanical loading by providing key signals that trigger bone modeling and remodeling; and (3) produce key factors and hormones that modulate functions of distant tissues (eg, FGF23 affecting phosphate excretion and $1,25(OH)_2D$ metabolism).

C. Osteoclast   The osteoclast is a multinucleated giant cell that is specialized for resorption of bone. Osteoclasts are terminally differentiated cells that arise from hematopoietic precursors in the monocyte lineage and do not divide. Osteoblasts stimulate both osteoclast formation and activation via the cell surface molecule RANKL and soluble RANKL released into the bone microenvironment. RANKL stimulates its cognate ligand RANK on the surface of osteoclast precursors and mature osteoclasts (Figure 8–22). Osteoblasts also elaborate macrophage colony stimulating factor-1 (M-CSF), which is required for the full effects of RANKL on osteoclastogenesis. In addition, osteoblasts and other cells produce a decoy receptor, osteoprotegerin (OPG), which binds to RANKL and blocks its actions. PTH and $1,25(OH)_2D$ both increase RANKL production by osteoblasts, as do many cytokines including IL-1, IL-6, and IL-11. TNF-alpha potentiates the ability of RANKL to stimulate osteoclastogenesis, whereas IFN-gamma blocks this process by direct effects on the osteoclast.

To resorb bone, the motile osteoclast alights on a bone surface and seals off an area by forming an adhesive ring. Having isolated an area of bone surface, the osteoclast develops above the surface an elaborately invaginated plasma membrane structure called the **ruffled border** (Figure 8–23). The ruffled border is a distinctive organelle, but it acts essentially as a huge lysosome that dissolves bone mineral by secreting acid onto the isolated bone surface and simultaneously breaks down bone matrix by secretion of proteases, in particular cathepsin K. The collagen breakdown products from bone resorption can be assayed in serum and urine as measures of bone resorption rates (*N*- and *C*-telopeptides). Bone resorption can be controlled in several ways: by regulating the formation of

FIGURE 8–22 Regulation of osteoclast formation and activation by the osteoblast. The osteoblast stimulates osteoclast formation and activation via the interaction of RANKL on the osteoblast with RANK on the osteoclast. M-CSF secreted by the osteoblast enhances the process via its receptor on the osteoclast C-FMS, whereas OPG, also secreted by the osteoblast, blocks the process by blocking RANKL. RANKL expression is stimulated by 1,25(OH)$_2$D and PTHrP/PTH as well as by the cytokines IL-1, IL-6, and IL-11.

osteoclasts via changes in cell numbers, by regulating the activity of mature osteoclasts, and by altering the function of key enzymes that degrade the bone matrix. The mature osteoclast has receptors for CT but does not appear to have PTH or vitamin D receptors.

BONE MODELING AND REMODELING

Bone remodeling is a continuous process of breakdown and renewal that occurs throughout life. During childhood and adolescence, remodeling proceeds at a vigorous rate but is quantitatively overwhelmed by the concomitant occurrence of bone modeling and linear growth. Once peak bone mass has been established, remodeling supervenes as the common mechanism by which bone mass is modified for the remainder of a person's life. Each remodeling event is carried out by individual "bone remodeling units" on bone surfaces throughout the skeleton (see Figure 8–21). Normally, about 90% of these surfaces lie dormant, covered by a thin layer of lining cells. Following physical or biochemical signals, precursor cells from the bone marrow migrate to specific loci on the bone surface, where they fuse into multinucleated bone-resorbing cells known as osteoclasts that then excavate a resorption cavity into the bone.

FIGURE 8–23 Osteoclast-mediated bone resorption. The osteoclast attaches to the bone surface via integrin-mediated binding to bone matrix bone proteins. When enough integrin binding has occurred, the osteoclast is anchored and a sealed space is formed. The repeatedly folded plasma membrane creates a ruffled border. Secreted into the sealed space are acid and enzymes forming an extracellular lysosome. (Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism*. 3rd ed. New York: McGraw-Hill Education; 1995.)

FIGURE 8–24 Schematic representation of the four principal stages involved in the formation of a new basic structural unit in cortical bone. (Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism.* 3rd ed. New York: McGraw-Hill Education; 1995.)

Cortical bone is remodeled from within by cutting cones, groups of osteoclasts that cut tunnels through the compact bone (Figure 8–24). They are followed by trailing osteoblasts that line the tunnels with a cylinder of new bone, progressively narrowing the tunnels until all that remains are the tiny haversian canals by which the cells left behind as resident osteocytes are fed. The packet of new bone formed by a single cutting cone is called an osteon (see Figure 8–20).

By contrast, trabecular resorption creates scalloped areas of the bone surface called Howship lacunae. Two to three months after initiation, the resorption phase reaches completion, having created a cavity about 60 μm deep. This is accompanied by ingress from marrow stroma into the base of the resorption cavity of precursors for bone-forming osteoblasts. These cells develop an osteoblastic phenotype, expressing characteristic bone-specific proteins such as alkaline phosphatase, osteopontin, and osteocalcin, and begin to replace the resorbed bone by elaborating new bone matrix. Once the newly formed osteoid reaches a thickness of about 20 μm, mineralization begins. Completion of a full remodeling cycle normally takes about 6 months (see Figure 8–21).

Bone remodeling does not absolutely require systemic hormones except to maintain intestinal absorption of minerals by $1,25(OH)_2D$ and thus ensure an adequate supply of calcium and phosphate. For example, bone is relatively normal, aside from being in a state of low turnover and being dense, in patients with hypoparathyroidism. However, systemic hormones use the "bone pool" as a source of minerals for regulation of extracellular calcium homeostasis. When they do, the coupling mechanism ensures that bone is replenished. For example, when bone resorption is activated by PTH to provide calcium to correct hypocalcemia, bone formation also increases, tending to replenish lost bone. Although

the role of the osteoblast in regulating osteoclast activity is reasonably well understood, the mechanism by which osteoblasts are recruited to sites of bone resorption is not. One possibility is that bone resorption releases IGF-1 from the bone matrix, and IGF-1 then stimulates osteoblast proliferation and differentiation. However, it is also becoming clear that the osteoclast itself can regulate osteoblast differentiation through bidirectional signaling molecules of the ephrin/Eph family. Transforming growth factor (TGF) beta is also released from the bone matrix with resorption, which also plays a role in stimulating formative responses.

If the replacement of resorbed bone matched the amount that was removed, remodeling would lead to no net change in bone mass. However, small deficits in bone mass persist on completion of each cycle, reflecting inefficiency in remodeling dynamics. Consequently, lifelong accumulation of remodeling deficits underlies the well-documented phenomenon of age-related bone loss, a process that begins shortly after growth stops. Alterations in remodeling activity represent the final pathway through which diverse stimuli, such as dietary insufficiency, hormones, and drugs affect bone balance. A change in whole body remodeling rate can be brought about through distinct perturbations in remodeling dynamics. Changes in the hormonal milieu often increase the activation of remodeling units. Examples include hyperthyroidism, hyperparathyroidism, and hypervitaminosis D. Other factors may impair osteoblastic functional adequacy, such as high doses of glucocorticoids or ethanol. Yet other perturbations, such as estrogen or androgen deficiency, may augment osteoclastic resorptive capacity. At any given time, a transient deficit in bone exists called the **remodeling space** representing sites of bone resorption that have not yet filled in. In response to any stimulus that alters the birth rate of new remodeling units, the remodeling space either increases or decreases accordingly until a new steady state is established, and this adjustment is seen as an increase or decrease in bone mass.

OSTEOPOROSIS

Osteoporosis is a condition of low bone mass and microarchitectural disruption that results in fractures with minimal trauma. The term **primary osteoporosis** denotes reduced bone mass and fractures in postmenopausal women (postmenopausal osteoporosis) or in older men and women due to age-related factors. The term **secondary osteoporosis** refers to bone loss resulting from specific clinical disorders, such as thyrotoxicosis, glucocorticoid excess and many others (Table 8–9). There is overlap in these designations; for example, many postmenopausal women with low bone mass have vitamin D insufficiency or deficiency which could be considered a secondary cause of osteoporosis.

Osteoporotic fractures are a major public health problem for older women and men in Western society. Half the men and women over age 55 have low bone mass (osteopenia) or osteoporosis, placing them at increased risk of fracture. At any age, women experience twice as many osteoporosis-related fractures as men, reflecting gender-related differences in skeletal properties as well as the almost universal loss of bone at menopause. However,

TABLE 8–9 Causes of secondary osteoporosis.

Endocrinopathies
Thyrotoxicosis
Hyperprolactinemia
Primary hyperparathyroidism
Acromegaly
Hypogonadism
Glucocorticoid excess
Gastrointestinal/nutritional
Vitamin D deficiency
Chronic liver disease
Celiac sprue
Malabsorption
Medications
Glucocorticoids
Androgen deprivation therapy with gonadotropin-releasing hormone agonists
Certain anticonvulsants
Excess thyroid hormone replacement
Other
Alcoholism
Osteogenesis imperfecta
Rheumatoid arthritis
Multiple myeloma
Chronic obstructive pulmonary disease
Idiopathic hypercalciuria

TABLE 8–10 Risk factors for fractures.

Advanced age
History of fracture after age 50
First-degree relative with a fracture
Current tobacco use
Low body weight
Inability to rise without using arms
Lifelong low calcium intake
Vitamin D deficiency
Inactivity/bed rest
Early estrogen loss
Low testosterone in men
Dementia
Alcoholism
Impaired vision
History of falls
Low bone mineral density

the number of osteoporotic fractures in older men is not trivial (Figure 8–25). One of six Caucasian women suffers a hip fracture; mortality after hip fracture is as high as 20% in the first year. One-third of hip fractures occur in men and have been associated with an even higher mortality rate than in women. Billions of dollars are spent annually for acute hospital care of hip fracture alone. The consequences of vertebral deformity are also significant and include chronic pain, inability to conduct daily activities, depression, increased mortality, and high risk of additional vertebral fractures.

Typically, fractures attributed to bone fragility are those due to trauma equal to or less than a fall from a standing position.
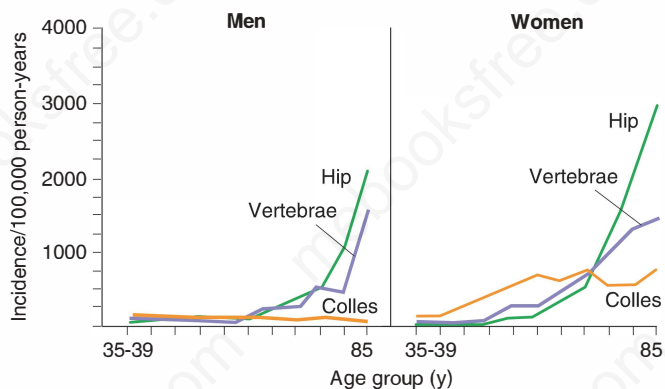
FIGURE 8–25 Incidence rates for the three common osteoporotic fractures (Colles, hip, and vertebral) in men and women, plotted as a function of age at the time of the fracture. (Reproduced with permission from Cooper C, Melton LJ. Epidemiology of osteoporosis. *Trends Endocrinol Metab.* 1992 Aug;3(6):224-229.)

Common sites of fragility-related fractures include the vertebral bodies, distal forearm, and proximal femur, but because the skeletons of patients with osteoporosis are diffusely fragile, other sites, such as ribs and long bones, also fracture with high frequency. Vertebral compression fractures are the most common fragility-related fractures. Pain sufficient to require medical attention occurs in approximately one-third of vertebral fractures, the majority being detected only when height loss or spinal deformity (kyphosis) occurs. However, even asymptomatic vertebral fractures confer an increased risk of future vertebral fractures and mortality. Thus, incidentally detected vertebral fractures have the same significance as a clinical vertebral fracture in this regard. Furthermore, even fractures that appear to be due to trauma portend an increased future fracture risk. Both skeletal and extraskeletal factors determine fracture risk (Table 8–10).

Gain, Maintenance, and Loss of Bone

The amount of bone mineral present at any time in adult life represents that which has been gained at skeletal maturity (peak bone mass) minus that which has been subsequently lost. Bone acquisition is completed in the late teenage years and early twenties in girls and by the second decade in boys. Heredity accounts for most of the variance in bone acquisition. African American adolescent girls have greater calcium retention on a fixed calcium intake compared to Caucasian adolescent girls, potentially explaining at least part of the racial differences in bone mass. Specific genes implicated in bone acquisition include those affecting body size, hormone responsiveness, and bone-specific proteins such as type 1 collagen and TGF beta.

The **Wnt pathway** plays a central role in determining bone mass. The low density lipoprotein (LDL) receptor-like protein 5/6 (LRP5/6) along with the frizzled protein serves as coreceptor for members of the Wnt family of ligands. Activation of Wnt and its downstream signaling pathways mediates increased osteoblastic activity (gene expression, cell proliferation and differentiation). Mutations that render LRP5 constitutively active are associated with high bone mass, and autosomal recessive inactivating mutations in *LRP5* play a causative role in the

osteoporosis-pseudoglioma syndrome. These mutations lead to accelerated bone loss and fractures as well as blindness. These two disorders of bone mass have firmly established Wnt, LRP5, and frizzled as essential molecules in osteoblast-mediated bone formation.

Other factors contributing to the acquisition of peak bone mass include circulating gonadal steroids, physical activity, and nutrient intake. In adolescents, calcium retention is maximal at a calcium intake of 1300 mg/d; this is the current recommended intake for individuals aged 9 to 18 years. Adolescence is a critical time for the skeleton; bone gained during adolescence accounts for about 60% of final adult bone mass. Assuming adequate exposure to key nutrients, physical activity, and reproductive hormones, adolescent growth supports acquisition of the maximum bone mass permitted by the genetic endowment. Estradiol plays a decisive role in the initiation of adolescent growth and bone acquisition. Rare examples have been reported of young men bearing mutations in the estradiol receptor or in aromatase, the enzyme that converts androgen to estrogen. Although these men had normal or increased circulating concentrations of testosterone, estradiol was either absent or totally ineffective. In all cases, severe deficits in BMD were observed, and these patients had not undergone the anticipated acceleration in linear growth at the time of puberty. In the patients with aromatase deficiency, substantial gains in bone were observed shortly after initiating estrogen therapy.

Adolescent bone acquisition falters in the face of inadequacies in diet, physical activity, or reproductive function, resulting in a lower peak bone mass and less reserve to accommodate future losses. Recent trends in habitual physical activity and calcium intake for North American teenagers, particularly girls, offer little encouragement in this regard. A list of representative conditions

TABLE 8–11 Examples of disorders associated with reduced peak bone mass.

Anorexia nervosa
Ankylosing spondylitis
Immobilization (bed rest, non-ambulatory status)
Cystic fibrosis
Delayed puberty
Exercise-associated amenorrhea
Galactosemia
Intestinal or renal disease
Marfan syndrome
Osteogenesis imperfecta

often associated with reduced peak bone mass acquisition is presented in Table 8–11.

Once peak bone mass is achieved, bone mass remains fairly stable until the late third or early fourth decades (Figure 8–26). Successful bone maintenance requires continued attention to the same factors that influenced bone acquisition: diet, physical activity, and sex hormone status. Maintenance of bone requires sufficiency in all areas, and one cannot compensate for a deficiency in another. For example, amenorrheic athletes lose bone despite frequent high-intensity physical activity and supplemental calcium intake. Successful bone maintenance is also jeopardized by known toxic exposures such as smoking, alcohol excess, and immobility as well as by systemic illnesses and many medications.

Bone Loss Associated with Estrogen Deficiency

Estrogen deprivation promotes bone remodeling by releasing constraints on the production and activity of cytokines, particularly
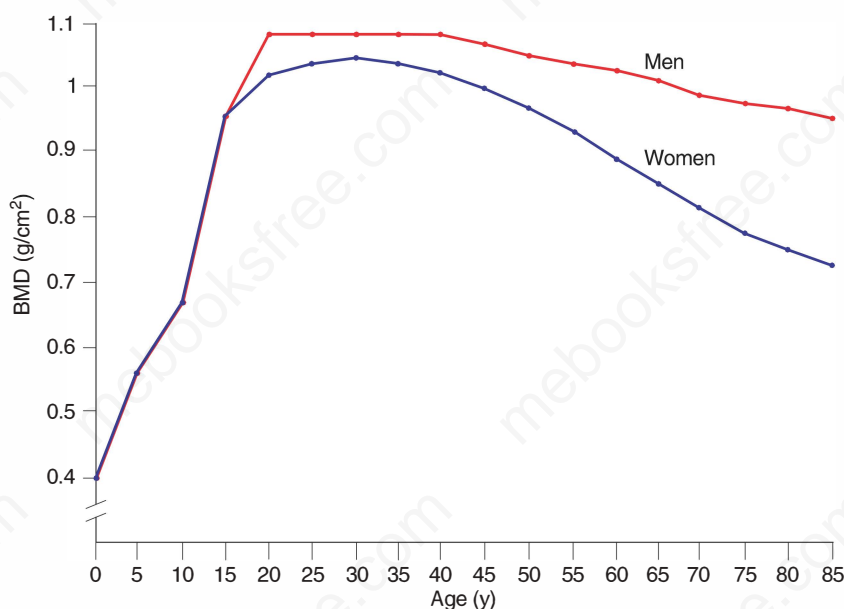
FIGURE 8–26 Mean bone mineral density by DXA in white males (upper curve) and females (lower curve) aged 5 to 85 years. (Data from Southard RN, et al. Bone mass in healthy children: measurement with quantitative DXA. *Radiology.* 1991;179:735; and from Kelly TL. Bone mineral reference databases for American men and women. *J Bone Miner Res.* 1990;5(Suppl 2):702.)

IL-6, which stimulates the proliferation of osteoclast precursors. Estradiol suppresses IL-6 secretion by marrow stromal osteoblastic cells. Treatment of mice with neutralizing IL-6 antibodies suppresses osteoclast production after oophorectomy, and *IL-6* gene knockout prevents bone loss in oophorectomized mice. Thus, a strong case implicates IL-6 as a critical molecule involved in estrogen action in bone. Subsidiary roles for IL-1 and IL-11 have also been described. As discussed earlier, the RANK-RANKL-OPG system acts as a bone regulatory system through which osteoblast signals stimulate the production of osteoclasts. With estrogen deficiency, OPG secretion is low, permitting a robust response of osteoclast precursors to RANKL, thereby increasing bone turnover.

With accelerated bone turnover, delivery of calcium from bone to the circulation increases, and the resultant subtle increase in plasma calcium concentration suppresses the secretion of PTH, thereby enhancing calciuria, suppressing renal production of $1,25(OH)_2D$, and reducing intestinal calcium absorption efficiency. At menopause, loss of endogenous estrogen promotes an increase in daily calcium loss from 20 mg to about 60 mg, reflecting a relative increase in bone resorption over formation activity. Although small, this magnitude of change in mineral balance accounts, after a decade, for about 13% of an original whole body calcium mass of 1000 g, equivalent to a standard deviation in BMD (see Figure 8–26), and leads to a two- to threefold increase in the risk of fracture. In the 5 years after menopause, women not on estrogen can lose as much as 5% to 8% of their bone mass. To accommodate menopausal changes in calcium economy by dietary means alone, a rise in daily calcium intake from 1000 mg to about 1500 mg would be necessary. Increased calcium intake can slow, but usually does not entirely prevent, the bone loss that occurs with menopause.

Bone Loss in Later Life

Progressive deficits in renal and intestinal function impair whole body calcium economy during normal human aging. These deficits include progressive inefficiency of vitamin D production by the skin as well as declining ability to convert 25(OH)D to $1,25(OH)_2D$ in the kidney. Consequently, intestinal calcium absorption becomes less efficient, leading to modest reductions in plasma ionized calcium activity and compensatory hypersecretion of PTH. PTH maintains blood calcium concentrations by activating new bone remodeling units, although as a result of its inherent inefficiency, increased bone remodeling leads to accelerated bone loss. Little can be done to counteract remodeling inefficiency, but the impact of these physiologic deficits can be minimized by suppressing the stimulus for PTH secretion by consuming adequate amounts of dietary calcium and vitamin D.

Diagnosis of Osteoporosis

The diagnosis of osteoporosis may be obvious in patients who have sustained fragility fractures (Figure 8–27), but noninvasive methods to estimate BMD are required to identify high-risk patients who have not yet sustained a fracture. Due to the significant consequences of fractures, diagnosis prior to fracture is essential. Several techniques have been developed for this purpose, but DXA offers the most precise measurements at multiple skeletal sites for the least amount of radiation exposure. Current
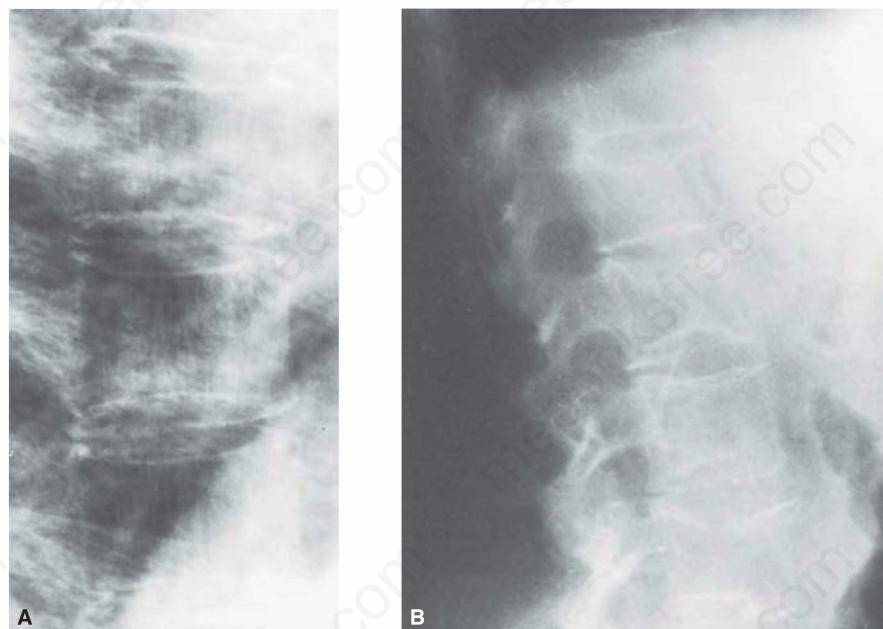
FIGURE 8–27   **A.** Magnified x-rays of thoracic vertebrae from a woman with osteoporosis. Note the relative prominence of vertical trabeculae and the absence of horizontal trabeculae. **B.** Lateral x-ray of the lumbar spine of a woman with postmenopausal osteoporosis. Note the increased density of the superior and inferior cortical margins of vertebrae, the marked demineralization of vertebral bodies, and the central compression of articular surfaces of vertebral bodies by intervertebral disks. (Used with permission from Dr. G. Gordan.)

guidelines recommend that all women 65 years and older be screened for osteoporosis using bone densitometry. The National Osteoporosis Foundation (NOF) Guideline recommends screening younger postmenopausal women with risk factors for fracture, while the United States Preventive Services Task Force recommends screening women under 65 years whose estimated fracture risk is at least that of a 65-year-old woman without additional risk factors. The NOF Guideline recommends screening men 70 years and older, and targeted screening of men ages 50 to 69 if there is a risk factor for fracture. There are no recommendations for routine screening of premenopausal women or men under 50 years, but younger men and women should have bone density testing if clinically indicated (eg, clinical evidence of bone fragility, chronic glucocorticoid therapy, and so forth).

BMD is a highly significant predictor of fracture risk. For each standard deviation BMD below the average BMD of individuals at the age of peak bone mass, fracture risk approximately doubles. Site-specific measurements best predict fracture risk at that site (ie, hip BMD best predicts hip fracture), but a measurement at any site predicts overall fracture risk. Based on data from Caucasian postmenopausal women and originally intended to be used as an epidemiologic tool, a World Health Organization (WHO) panel in 1994 offered an absolute standard for categorization of femoral neck BMD, and the International Society for Clinical Densitometry then extended the criterion to total hip and total posteroanterior lumbar spine BMD. By this standard, BMD values 2.5 or more SD below the average value for a person at the age of peak bone mass (a T-score of –2.5 or below) are categorized as **osteoporosis**. BMD values between –1.0 and –2.5 SD below the standard (T-score between –1 and –2.5) are categorized as low bone mass or **osteopenia**.

While not their original intended purpose, these T-score cut points became integrated into clinical practice and have been used to place individuals with low BMD into diagnostic categories. Problematically, there are no risk thresholds at these values; an individual with a T-score of –2.4 and an individual with a T-score of –2.5 have nearly identical fracture risk, but one is considered to have osteoporosis and one is not. Additionally, two individuals with the same T-score value can have very different fracture risk. A woman in her fifties with no personal or family history of fracture but a femoral neck T-score of –2.5 has a 10-year risk of fracture of less than 10%. However, an 80-year-old woman with the same femoral neck T-score of –2.5, whose mother had a hip fracture and who has a personal history of fracture, has a 10-year risk of fracture of more than 50%. It has long been recognized that the combination of risk factors for fracture and BMD can improve fracture risk prediction. The WHO Fracture Risk Assessment tool (http://www.shef.ac.uk/FRAX) formally estimates 10-year fracture risk from femoral neck bone density and clinical risk factors. The FRAX tool has greatly improved the clinician's ability to estimate fracture risk for an individual patient. The levels of risk at which treatment should be initiated are country-specific, but a cost-effectiveness analysis in the United States suggests that individuals with a 10-year probability of major osteoporotic fracture of 20% or higher or of hip fracture of 3% or higher should be considered for treatment. Limitations of this fracture prediction model include

that it only incorporates BMD by DXA at the femoral neck, many risk factors are not available in the model, and there is no way to incorporate the effect of prior treatment on fracture risk.

For premenopausal women and men under 50 years, assessment of BMD should be based on age-matched normative data (Z-scores). Although there are no formal guidelines and a paucity of long-term data, most experts consider a BMD more than two standard deviations below the age- and gender-matched normal reference population (a Z score of –2 or lower) to indicate low BMD in a young person. However, BMD in younger individuals must be interpreted in light of the clinical situation and other risk factors for fracture.

MANAGEMENT OF OSTEOPOROSIS

Nonpharmacologic Aspects of Osteoporosis Management

Much of the nonpharmacologic management of osteoporosis involves issues that are not specific for bone. These include exercise, fall prevention, and nutrition, and as appropriate, pain control, depression therapy, and assisting patients to organize activities of daily living.

A. Exercise   Minimal attention has been devoted to the role of exercise for patients with established osteoporosis. Patients and physicians often show reluctance to participate in exercise because of concerns for additional injury. However, activity avoidance aggravates bone loss and places the skeleton in even greater jeopardy, whereas improving muscle strength, particularly of the back extensor groups, constitutes a powerful means for reducing pain and increasing functional capacity. Even the most frail individuals can successfully implement exercise regimens with the assistance of a trained physical therapist. Exercise regimens that include a component of weight bearing and/or resistance exercise have been shown to increase BMD approximately 1% per year in randomized trials. If exercise is discontinued, these gains in BMD are lost.

Because the great majority of hip fractures are the immediate result of a fall, strategies aimed at reduction of falls are of critical importance. Muscle weakness is an important predictor of fall risk, and decreased muscle mass and strength are consequences of normal human aging. Resistance exercise (ie, weight training) promotes muscle strength even in very old men and women, and evidence suggests that increased lower extremity strength may reduce falls risk by improving postural stability. Thus, a widely disseminated program of leg-strengthening exercise could lower the risk of falling and reduce hip fracture incidence, even if no changes in BMD are achieved.

B. Other fall prevention measures   Proper footwear and installation of safety features around the home may minimize the risk of falling. Such features include bathroom safety rails, night lights, and rails and lighting for stairways. Elimination of floor clutter is important. Canes and other walking aids should

be recommended to patients with an unsteady gait. Physicians may need to overcome substantial patient resistance, but these aids can be lifesaving. Visual acuity should be maximized, and visual impairments corrected to the extent possible. Attention should be given to other comorbid conditions that may affect fracture risk, such as pain and depression, and the effects of medications that may cause diminished alertness and responsiveness by causing sedation or orthostatic hypotension.

C. Calcium   In the seventh decade and beyond, supplemental calcium and vitamin D therapy decrease bone turnover, increase bone mass, and have been shown in clinical trials to decrease fracture incidence. Studies confirming the clinical efficacy of various pharmacologic agents to increase BMD or decrease fracture have been performed in the setting of calcium and vitamin D sufficiency. Thus, adequate intake of calcium and vitamin D should be considered a basic approach for prevention and treatment of osteoporosis in all patients (Table 8–12).

The current recommended intake of calcium is 1300 mg/d for adolescents, 1000 mg/d for adult women up to age 50 and men up to age 70, and 1200 mg/d for women over age 50 and men over age 70. A dairy-free diet contains approximately 250 mg/d of calcium. Achieving the recommended calcium intake goals by food alone typically requires consumption of three to four servings per day of dairy foods or calcium-fortified foods. Patients who are unable to increase dietary calcium by diet alone may choose from many palatable, low-cost calcium preparations, the most frequently prescribed being the carbonate salt. Others include the lactate, gluconate, and citrate salts. Absorption of most commonly prescribed calcium products is similar. For many patients, cost and palatability outweigh modest differences in absorption. Calcium carbonate requires an acid environment for optimal absorption. For older individuals who may have hypo- or achlorhydria, taking calcium carbonate tablets with meals generally provides adequate acidity for this requirement. To maximize absorption, calcium should be taken in doses of 500 mg or less at a time and spread throughout the day. Calcium can interfere with the absorption of iron and L-thyroxine and should be ingested separately from these medications. There is no evidence that calcium intake beyond the current recommended levels results in additional benefits, and excessive supplemental intake modestly increases risk of nephrolithiasis. Concern has been raised about potential risk of cardiovascular events with calcium supplement use, although reports are inconsistent and this topic is controversial. A sensible approach is to encourage calcium from food sources and to use supplements only when necessary to attain the recommended total intake levels.

D. Vitamin D   In recent years, it has become necessary to reassess our concepts of vitamin D sufficiency (see also Vitamin D discussed earlier). Although circulating 25(OH)D concentrations above 10 ng/mL appear adequate for preventing histologic evidence of osteomalacia, a case has been made that values below 25 to 30 ng/mL may be associated with hypersecretion of PTH, proximal muscle weakness predisposing to falls, and increased bone turnover. In some regions of the United States, healthy adults show average 25(OH)D concentrations of 30 ng/mL or above, but in others, such as New England or in states farther north than Missouri, the ultraviolet component of sunlight fails to contain sufficient amounts, at the critical region of approximately 280 to 320 nm, to support cutaneous vitamin D synthesis. In those areas, vitamin D insufficiency is very common, particularly in frail, homebound, or hospitalized patients. Vitamin D supplementation in individuals with marginal or deficient vitamin D status improves intestinal calcium absorption, suppresses PTH and bone remodeling, and increases bone mass. In combination with calcium, it reduces fracture risk. Daily vitamin D supplementation may also reduce the risk of falls.

Pharmacologic Approaches to Osteoporosis Management

Drugs used for prevention and treatment of osteoporosis act either by decreasing the rate of bone resorption, thereby slowing the rate of bone loss, or by increasing bone formation (Table 8–13). Because of the coupling of bone resorption and formation, agents that decrease bone resorption ultimately decrease the rate of bone formation. Thus, increases in BMD, representing a reduction of

TABLE 8–12 Current recommendations for adequate daily intake of calcium and food sources of calcium.

Children/Adolescents	
Age 1-3	700 mg
Age 4-8	1000 mg
Ages 9-18	1300 mg
Adults	
Age 19-50	1000 mg
Age 51-70 (men)	1000 mg
Age 51-70 (women)	1200 mg
Age >70	1200 mg
Food Sources of Calcium	
Dairy-free diet	250 mg
Milk (8 oz)	300 mg
Calcium set tofu (1/2 cup)	400 mg
Cheese (1 oz)	250 mg
Yogurt (1 cup)	300 mg
Cooked greens (1 cup)	100 mg
Fortified foods	Varies

TABLE 8–13 Treatment for osteoporosis.

Antiresorptive Agents
Estrogen
SERMs
Calcitonin
Bisphosphonates
RANK-ligand inhibitor (denosumab)
Bone-Forming Agents
Parathyroid hormone
Strontium ranelate[a]

[a]Not approved in the United States.

the remodeling space to a new steady-state level, are commonly observed during the first year or two of therapy, after which BMD increases slowly and may even reach a plateau. It is not uncommon for there to be minimal changes in BMD observed with treatment in clinical practice. Analysis of data from the Fracture Intervention Trial demonstrated that women treated with alendronate whose bone density did not increase during treatment still benefited, with rates of fracture reduction similar to the women whose BMD measurably increased. Thus, women on treatment with stable BMD should not be viewed as nonresponders or treatment failures.

Antiresorptive Agents

A. Estrogen   Conjugated equine estrogens (0.625 mg/d) increase BMD and reduce the risk of hip and vertebral fractures by about 35%. The Women's Health Initiative (WHI) demonstrated a lack of cardiovascular disease prevention in estrogen-treated postmenopausal woman and small but statistically significant increases in breast cancer and venous thromboembolic phenomena with estrogen and progesterone therapy (hormone therapy; HT) in postmenopausal women. These data led to a marked change in prescribing of HT for postmenopausal women such that only patients suffering from significant vasomotor effects of estrogen deficiency and early in menopause are now prescribed HT and for the shortest possible time. Reduced BMD is not considered an indication for HT, because effective alternative therapies for skeletal preservation exist. Low dose conjugated estrogens and estradiol increase BMD, but low dose therapy has not been tested for antifracture effects. Tibolone, which has estrogenic actions, is approved for osteoporosis treatment outside of the United States.

B. Selective estrogen response modulators   SERMs are molecules that act as estrogens on some tissues but antiestrogens on others. Raloxifene, the only SERM currently approved by the United States FDA for osteoporosis treatment, is an estrogen agonist at bone and liver, promoting conservation of BMD and lowering of LDL cholesterol concentrations. Raloxifene has been shown to decrease the vertebral fracture incidence in older osteoporotic women. It is inert at the endometrium and is a potent antiestrogen at the breast, decreasing breast cancer risk among high risk women. Side-effects include venous thromboembolism and hot flashes. Bazedoxifene, another SERM, is FDA-approved for the prevention (but not treatment) of osteoporosis in combination with estrogen.

C. Calcitonin   The role of endogenous calcitonin in calcium metabolism is unclear. However, given as a pharmacologic agent, calcitonin inhibits osteoclastic bone resorption. Intranasal and injectable formulations are approved for postmenopausal osteoporosis treatment. CT nasal spray (200 IU/d) increases BMD by 2% to 3% and reduces vertebral fractures by about 35%. CT has not been shown to reduce hip or nonspine fractures in randomized trials. CT may provide some analgesic benefit in patients with painful vertebral fractures, possibly through central nervous system effects. Side-effects of nasal spray CT appear mild and include rhinitis. Injectable CT may cause flushing, nausea, and local injection reactions, making the nasal option the preferred form of administration, but both formulations are used infrequently now that more potent agents are available.

D. Bisphosphonates   Bisphosphonates bind avidly to the hydroxyapatite crystals in bone, particularly at sites of active remodeling. The currently approved aminobisphosphonates interfere with protein prenylation in the osteoclast, impairing osteoclast-mediated bone resorption as well as enhancing osteoclast apoptosis. Four bisphosphonates, alendronate, risedronate, ibandronate, and zoledronic acid are FDA-approved for prevention and treatment of osteoporosis. Three of them (alendronate, risedronate, and zoledronic acid) are approved for treatment of glucocorticoid-induced osteoporosis in men and women and for the treatment of osteoporosis in men.

All the approved bisphosphonates have been shown to reduce vertebral fractures by approximately 50% to 60%. Alendronate, risedronate, and zoledronic acid have also been shown to reduce the risk of nonvertebral fractures. In clinical trials, hip fractures were reduced by 51% after treatment with alendronate, 30% after treatment with risedronate, and 41% after annual infusions of zoledronic acid for 3 years. Head-to-head comparisons, however, have not been performed. Treatment of individuals (men and women) who have had a hip fracture with zoledronic acid has also been shown to reduce subsequent fractures and decrease mortality in individuals. Ibandronate has not been shown to reduce the risk of nonvertebral fractures except in the post hoc analysis of women with T-scores less than −3. The fracture trials for all the bisphosphonates were conducted using daily dosing, except for zoledronic acid, which was given as a 5 mg annual intravenous dose. Based on equivalent effects on BMD to daily dosing, alendronate and risedronate are approved for weekly dosing, and risedronate and ibandronate are approved for monthly dosing. Ibandronate is also approved for quarterly intravenous dosing.

Oral bisphosphonate absorption is very low, less than 1% of the administered dose. Consequently, patients must take these medications on first arising in the morning, on an empty stomach with the exception of plain water, and ingest nothing else for 30 minutes for alendronate and risedronate, and 60 minutes for ibandronate. Side-effects from the oral bisphosphonates are primarily upper gastrointestinal, most notably esophageal irritation if the pill does not clear the esophagus. Hence, they must be taken with at least 8 oz of water, and a patient must remain upright to ensure passage of the pill into the stomach, and oral bisphosphonates are contraindicated in patients with clinically significant esophageal disease. Other bisphosphonate-related side-effects include bone pain and, for the intravenous agents, an acute phase reaction temporally related to dosing and typically limited to the patient's first dose. Bisphosphonate administration may precipitate hypocalcemia in patients with low levels of 25(OH)D. None of the bisphosphonates are recommended for use in patients with significant renal impairment (creatinine clearance <35 mL/min).

Two rare but serious side-effects have been reported with bisphosphonate therapy. Osteonecrosis of the jaw is an area of exposed bone in the maxillofacial region that did not heal within 8 weeks of identification by a health-care provider. The vast majority of cases have occurred after dental work in patients with malignancies on intravenous bisphosphonates at more frequent intervals and at much higher doses of drug than the annual infusions for patients with osteoporosis. Several cases, however, have been reported in patients with osteoporosis. Estimates of the incidence of this complication related to oral bisphosphonate treatment for osteoporosis appear to be between 1 in 10,000 and less than 1 in 100,000 patient-treatment years. Rates are significantly higher in malignant conditions such as multiple myeloma or breast cancer. Osteonecrosis of the jaw is a difficult disorder to treat, and suspected cases should be referred to an experienced dental professional. The other rare side-effect is the atypical femur fracture. This is defined as a fracture in the shaft or subtrochanteric region of the femur that originates at the lateral cortex and has a transverse orientation, is noncomminuted or minimally comminuted, shows lateral cortical thickening, and occurs with minimal or no trauma. Although the relative risk of patients with atypical femur fractures taking bisphosphonates is high, the absolute risk of these fractures with bisphosphonates is low. Long-term use of bisphosphonates may be associated with higher risk.

Duration of bisphosphonate therapy is a topic of great interest, but limited data exist about long-term use. One potential approach, based on limited evidence, is to reassess a patient's fracture risk after 5 years of oral or 3 years of intravenous bisphosphonate therapy. For those at high risk for fracture, continued therapy might be appropriate, while for those not at high risk, a drug holiday might be considered.

E. RANKL inhibition   Denosumab, a human monoclonal IgG2 antibody to RANKL, has been shown to reduce spine fractures by 68%, hip fractures by 40%, and nonvertebral fractures by 20% in postmenopausal women with low bone density (T-score <−2.5 and >−4.0). In the pivotal trial establishing this antifracture efficacy, 60 mg denosumab was administered subcutaneously every 6 months for 3 years. Denosumab has been approved by the FDA for the treatment of postmenopausal women and men with osteoporosis at high risk for fracture. Denosumab is also FDA-approved for the treatment of men at high risk for fracture receiving androgen deprivation therapy for nonmetastatic prostate cancer and of women at high risk for fracture receiving adjuvant aromatase inhibitor therapy for breast cancer. Adverse events include eczema and cellulitis. Rare cases of osteonecrosis of the jaw and atypical femur fractures have been reported with denosumab therapy.

Bone-Forming Agents

A. Parathyroid hormone   Fractures, skeletal deformity, and bone pain are well-described manifestations of bone disease associated with severe primary hyperparathyroidism. It may seem counterintuitive, therefore, that administration of PTH to individuals with low BMD not only improves BMD but also reduces the risk of fractures. Yet, this is the case. The decisive element that determines the effects of PTH to be destructive or therapeutic appears to be whether PTH concentrations in blood are elevated constantly or only intermittently. In severe hyperparathyroidism, PTH concentrations remain high, varying little throughout the day. Administration of PTH as daily pulse injections has been shown to increase BMD in both animals and humans, and abundant evidence now indicates that PTH is truly an osteotropic agent. In a large clinical trial conducted in postmenopausal women with osteoporosis, recombinant human PTH(1-34) (teriparatide), given in daily subcutaneous doses for approximately 21 months, reduced the incidence of vertebral fractures by more than 60% and that of nonvertebral fractures by 50%. During the course of this trial, a long-term carcinogenicity study showed that rats given very high daily injections of rhPTH(1-34) showed a dose-related incidence of osteosarcoma. Clinical studies with this agent were immediately suspended, until an independent group of cancer biologists concluded that the rat osteosarcoma finding was not likely to predict a relationship to osteosarcoma in humans, at which time human trials were resumed. Teriparatide is approved by the FDA for use in both men and postmenopausal women with osteoporosis who are at high risk for fractures as well as in glucocorticoid-induced osteoporosis. Therapeutic use is limited to 2 years and is recommended for individuals who have had a fracture related to osteoporosis, who have multiple risk factors for fracture, or who cannot use other osteoporosis treatments. Side-effects of PTH therapy include hypercalcemia, hypercalciuria, and bone pain, although they are infrequent. PTH therapy is contraindicated for patients with conditions placing them at increased risk for osteosarcoma such as Paget disease of bone, unexplained elevations of alkaline phosphatase, open epiphyses, or patients who have received radiation therapy involving the skeleton. The very few reported cases of possible osteosarcoma approximate the background risk of osteosarcoma in the general population. After cessation of PTH therapy, bone loss resumes, and most if not all bone gained may be lost. It is strongly recommended that PTH therapy be followed with an antiresorptive agent, thus maintaining the gain in BMD achieved with PTH administration.

B. Other anabolic agents   PTH is the only anabolic agent approved by the FDA at this time, although other anabolic agents are under investigation. Outside the United States, strontium ranelate is used to treat osteoporosis.

GLUCOCORTICOID-INDUCED OSTEOPOROSIS

In children, chronic exposure to glucocorticoids impairs skeletal growth. Glucocorticoid-induced deficits in childhood bone mass reflect failure of normal skeletal acquisition occasionally compounded by bone loss. In adults, initiation of glucocorticoid therapy leads to rapid bone loss, resulting in significant BMD

deficits within even a few months with a greater impact on trabecular than cortical bone. The magnitude of bone loss can be large, typically approaching 40% reduction of the initial BMD. Approximately half of long-term glucocorticoid-treated patients suffer a fracture. Skeletal fragility is not the only risk factor for fracture in steroid-treated patients, because glucocorticoids (and the illnesses for which they are prescribed) are also associated with muscle weakness, which itself creates instability and increases the risk of falling.

Glucocorticoid-induced bone loss generally follows sustained administration of systemic doses greater than 5 mg/d of prednisone or its equivalent. Short, infrequent courses of glucocorticoids, such as for poison oak dermatitis, do not appear to have long-term skeletal consequences. Used according to manufacturer recommendations, most steroid inhalers appear to be without skeletal consequence, because systemic absorption of glucocorticoids is minimal. However, excessive use of standard preparations may achieve systemic concentrations, adequate to suppress the hypothalamic-pituitary axis and predictably lead to loss of bone mass. Patients receiving maintenance hydrocortisone dosages (eg, 20 mg/d) for adrenocortical insufficiency do not generally have an increased skeletal risk. However, based on accurate measurements of steroid production rates in normal humans, maintenance doses of hydrocortisone are now estimated to be lower (ie, 15 mg/d) than have traditionally been recommended (about 30 mg/d). Patients who continue to be treated according to older recommendations may be at risk of excessive bone loss.

Pathophysiology

Glucocorticoids affect mineral balance through alterations in renal, intestinal, and skeletal function.

A. Renal calcium losses   Within a few days after initiating therapy, direct inhibition of renal tubular calcium reabsorption leads to hypercalciuria, which is magnified by excessive dietary sodium intake and attenuated by thiazide diuretics.

B. Intestinal calcium losses   Glucocorticoids, given in high doses for several weeks, inhibit intestinal calcium absorption, lowering plasma ionized $[Ca^{2+}]$. This is thought to stimulate compensatory hypersecretion of PTH, which restores serum ionized $[Ca^{2+}]$ by increasing the efficiency of renal calcium reabsorption and by activating bone turnover to increase delivery of calcium to the circulation from bone. However, it has never been shown consistently that plasma PTH concentrations are actually elevated in steroid-treated patients. Evidence suggests that PTH action in bone and kidney may be enhanced even if there are minimal or no increases in detectable PTH. PTH may contribute to the pathophysiology of bone loss in glucocorticoid-treated patients. Consistent effects of glucocorticoids on the production, clearance, and circulating concentrations of vitamin D metabolites have not been demonstrable, so it appears that the intestinal actions of glucocorticoids are independent of the vitamin D system.

C. Skeletal losses   Glucocorticoids affect the skeleton by multiple mechanisms. There is evidence for direct stimulation of bone resorption via increased RANKL expression and decreased OPG production. Glucocorticoids also inhibit osteoblastic maturation and activity. Recent information indicates that glucocorticoids shift multipotent stem cell maturation away from osteoblastic lineage toward other cell lines, particularly adipocytes. Glucocorticoids also promote apoptosis in osteoblasts and osteocytes. In patients chronically treated with glucocorticoids, bone formation is suppressed. High-dose glucocorticoid administration also suppresses gonadotropin secretion and creates a hypogonadal state in both men and women. Thus, a component of the bone loss in some steroid-treated patients is likely to be related to loss of gonadal function.

Prevention and Treatment of Glucocorticoid-Related Osteoporosis

A. Reduction of steroid dose   Prevention of bone loss remains the approach most likely to give a favorable outcome. The most important aspect of a preventive strategy is to limit exposure to glucocorticoids and to find alternative treatments if possible. Opportunities should be sought for initiating dose reduction or switching to nonsystemic forms of steroid administration. Unfortunately, the use of alternate-day steroids, which is known to protect growth in steroid-treated children, appears not to offer skeletal protection to adults. The possibility that a bone-sparing glucocorticoid might be developed has been under consideration for many years, but no such agent is currently available. In patients who require immunosuppression following organ transplantation, the use of cyclosporine and tacrolimus has permitted reductions in glucocorticoid use. However, these other drugs also have been associated with lower BMD; whether long-term therapy will have less negative skeletal effects than glucocorticoids is not established.

B. Calcium and vitamin D   Providing supplemental calcium and vitamin D to glucocorticoid-treated patients normalizes serum ionized $[Ca^{2+}]$, suppresses PTH secretion, and reduces bone remodeling, thereby reducing the number of active remodeling units at a given time. So long as the patient continues to take glucocorticoids, however, suppression of osteoblast function does not reverse. Patients should receive 1200 mg/d of supplemental calcium on the condition that such supplementation does not result in hypercalciuria. Vitamin D dosing should be targeted to maintain "therapeutic" 25(OH)D levels. However, periodic surveillance of glucocorticoid-treated patients for urinary calcium excretion is warranted, because they are at higher risk of developing kidney stones as a consequence of hypercalciuria.

C. Physical activity   Skeletal immobilization precipitates bone loss, so inactivity must be avoided in steroid-treated patients. Some conditions such as polymyalgia rheumatica respond exuberantly to steroid therapy, giving patients a rapid

and complete recovery of function. These patients may have little difficulty maintaining a vigorous schedule of weight-bearing activities during treatment. Patients with other conditions may not regain full mobility and may have residual functional disabilities. It is particularly important that such patients receive physical therapy or other appropriate support for maintaining and restoring functional capacity.

D. Hormone therapy   Glucocorticoids decrease sex steroid production, so evaluation of gonadal hormone status should be considered in all glucocorticoid-treated patients. In some rheumatologic conditions (eg, systemic lupus erythematosus), estrogens have traditionally been avoided. If there are no contraindications, then estrogen/progestin replacement may be considered in premenopausal women with amenorrhea and hypoestrogenism, and testosterone should be considered for men with hypogonadism.

Pharmacologic Therapy of Glucocorticoid-Related Osteoporosis

Alendronate, risedronate, zoledronic acid, and teriparatide have been shown to increase BMD and decrease the risk of new vertebral fractures by up to 70% in patients on glucocorticoid therapy.

Because the bone loss associated with glucocorticoid therapy has a very rapid onset and occurs in approximately 50% of treated patients, starting a bisphosphonate or teriparatide along with the glucocorticoid therapy may be appropriate. It is prudent to assess BMD and to estimate fracture risk in patients beginning glucocorticoid therapy. Whether to initiate an antiosteoporosis medication and which agent to begin may then be decided based on baseline fracture risk, glucocorticoid dose, and anticipated duration of glucocorticoid therapy. Serial BMD assessment is warranted, with the frequency of testing influenced by whether or not osteoporosis therapy has already been initiated, the results of previous BMD testing, and glucocorticoid dose.

OSTEOMALACIA AND RICKETS

Osteomalacia and rickets are caused by the abnormal mineralization of bone and cartilage. Osteomalacia is a bone defect occurring after the epiphyseal plates have closed (ie, in adults). Rickets occurs in growing bone (ie, in children). Abnormal mineralization in growing bone affects the transformation of cartilage into bone at the zone of provisional calcification. As a result, an enormous profusion of disorganized, nonmineralized, degenerating cartilage appears in this region, leading to widening of the epiphyseal plate (observed radiologically as a widened radiolucent zone) with flaring or cupping and irregularity of the epiphyseal-metaphyseal junctions. This latter problem gives rise to the clinically obvious beaded swellings along the costochondral junctions known as the **rachitic rosary** and the swelling at the ends of the long bones. Growth is retarded by the failure to make new bone. Once bone growth has ceased (ie, after closure of the epiphyseal plates), the clinical evidence for defective mineralization becomes more subtle, and special diagnostic procedures may be required for its detection.

Pathogenesis

The best-known cause of abnormal bone mineralization is vitamin D deficiency (discussed earlier). Vitamin D, through its biologically active metabolites, ensures that the calcium and phosphate concentrations in the extracellular milieu are adequate for mineralization. Vitamin D may also permit osteoblasts to produce a bone matrix that can be mineralized and then allows them to mineralize that matrix normally. Phosphate deficiency can also cause defective mineralization, as in diseases in which phosphate is lost in the urine or poorly absorbed in the intestine. A number of hypophosphatemic conditions result from excess circulating FGF23 that can both decrease 1-hydroxylase activity and renal phosphate reabsorption. Phosphate deficiency may act independently or in conjunction with other predisposing abnormalities, because most hypophosphatemic disorders associated with osteomalacia or rickets also affect the vitamin D endocrine system. Dietary calcium deficiency has been shown to lead to rickets in children in the absence of vitamin D deficiency and may contribute to the osteomalacia of elderly adults who are also susceptible to vitamin D deficiency.

Osteomalacia or rickets may develop despite adequate levels of calcium, phosphate, and vitamin D if the bone matrix cannot undergo normal mineralization as a result of enzyme deficiencies, such as decreased alkaline phosphatase in patients with hypophosphatasia, or in the presence of inhibitors of mineralization, such as aluminum, fluoride, or etidronate. Table 8–14 lists diseases associated with osteomalacia or rickets according to their presumed mechanism. Several diseases appear under several headings, indicating that they contribute to bone disease by several mechanisms.

Diagnosis

The following discussion refers primarily to vitamin D deficiency. In children, the presentation of rickets is generally obvious from a combination of clinical and radiologic evidence. The diagnostic challenge is to determine the cause. In adults, the clinical, radiologic, and biochemical evidence of osteomalacia is often subtle. In situations in which osteomalacia should be suspected such as malnutrition or malabsorption, BMD is strikingly low, and the clinician must decide whether to obtain a bone biopsy for histomorphometric examination. This decision rests on the availability of resources to obtain and examine the biopsy specimen, the index of suspicion coupled with the lack of certainty from other diagnostic procedures, and the degree to which the therapeutic approach will be altered by the additional information. In many cases, a therapeutic trial of vitamin D supplements suffices to establish the diagnosis without the need for bone biopsy.

Clinical Features

A. Symptoms and signs   The clinical presentation of rickets depends on the age of the patient and, to some extent, the

TABLE 8–14 Causes of osteomalacia.

Disorders of the Vitamin D Endocrine System
Decreased bioavailability
Insufficient sunlight exposure
Nutritional vitamin D deficiency
Nephrotic syndrome (urinary loss)
Malabsorption (fecal loss)
Billroth type II gastrectomy
Sprue
Regional enteritis
Bariatric surgery
Pancreatic insufficiency
Cholestatic disorders
Cholestyramine
Abnormal metabolism
Liver disease
Chronic renal failure
Pseudovitamin D deficiency rickets
Tumor-induced osteomalacia
X-linked hypophosphatemia
Autosomal dominant hypophosphatemic rickets
Chronic acidosis
Anticonvulsants
Abnormal target tissue response
Hereditary vitamin D deficient rickets
Gastrointestinal disorders
Disorders of Phosphate Homeostasis
Decreased intestinal absorption
Malnutrition
Malabsorption
Antacids containing aluminum hydroxide
Diarrhea
Increased renal loss
X-linked hypophosphatemic rickets
Autosomal dominant hypophosphatemic rickets
Autosomal recessive hypophosphatemic rickets
Mutations in renal phosphate transporters (SLC34A1, SLC34A3, SLC20A2)
Tumor-induced osteomalacia
De Toni-Debré-Fanconi syndrome
Calcium Deficiency
Primary Disorders of Bone Matrix
Hypophosphatasia
Fibrogenesis imperfecta ossium
Axial osteomalacia
Inhibitors of Mineralization
Aluminum
Chronic renal failure
Total parenteral nutrition
Etidronate
Fluoride

FIGURE 8–28 The clinical appearance of a young child with rickets. The most striking abnormalities are the bowing of the legs and protuberant abdomen. Flaring of the ends of the long bones can also be appreciated. (Used with permission from Dr. Sara Arnaud.)

cause of the syndrome. The affected infant or young child may be apathetic, listless, weak, hypotonic, and growing poorly. A soft, somewhat misshapen head, with widened sutures and frontal bossing, may be observed. Eruption of teeth may be delayed, and teeth that do appear may be pitted and poorly mineralized. The enlargement and cupping of the costochondral junctions produce the rachitic rosary on the thorax. The tug of the diaphragm against the softened lower ribs may produce an indentation at the point of insertion of the diaphragm (Harrison groove).

Muscle hypotonia can result in a pronounced potbelly and a waddling gait. The limbs may become bowed, and joints may swell because of flaring at the ends of the long bones (including phalanges and metacarpals) (Figure 8–28). Pathologic fractures may occur in patients with florid rickets. After the epiphyses have closed, the clinical signs of rickets or osteomalacia are subtle and cannot be relied on to make the diagnosis. Patients with severe osteomalacia complain of bone pain and proximal muscle weakness. Difficulty climbing stairs or rising from chairs may be reported and should be looked for. Such individuals may have a history of fractures and be diagnosed as having osteoporosis.

B. Laboratory findings

1. **Biochemistry**—Vitamin D deficiency results in decreased intestinal absorption of calcium and phosphate. In conjunction with the resulting secondary hyperparathyroidism, vitamin D deficiency leads to an increase in bone resorption, increased excretion of urinary phosphate, and increased renal tubular reabsorption of calcium. The net result tends to be low-normal serum calcium, low serum phosphate, elevated serum alkaline phosphatase, increased plasma PTH, decreased urinary calcium, and increased urinary phosphate levels. Finding a low serum 25(OH)D level, in combination with these other biochemical alterations, strengthens the diagnosis of vitamin D deficiency. The serum 1,25(OH)$_2$D level may be normal, making this determination less useful for the diagnosis of osteomalacia. Both 25(OH)D and 1,25(OH)$_2$D levels may be reduced in patients with liver disease or nephrotic syndrome because the binding proteins for the vitamin D metabolites are low

secondary to decreased production (liver disease) or increased renal losses (nephrotic syndrome; discussed later). Such individuals may have normal free concentrations of these metabolites and so are not vitamin D-deficient. Other factors, such as age and diet, must be considered. For example, serum phosphate values are normally lower in adults than in children. Dietary history is important, because urinary phosphate excretion and, to a lesser extent, urinary calcium excretion reflect dietary phosphate and calcium content. Because phosphate excretion depends on the filtered load (the product of the glomerular filtration rate and plasma phosphate concentration), urinary phosphate levels may be reduced despite the presence of hyperparathyroidism, when serum phosphate levels are particularly low. Expressions of renal phosphate clearance that account for these variables (eg, renal threshold for phosphate, or TmP/GFR) are a better indicator of renal phosphate handling than is total phosphate excretion. The TmP/GFR can be calculated from a nomogram using measurements of a fasting serum and urine phosphate concentration.

2. **Histologic examination**—Transcortical bone biopsy is the definitive means of making the diagnosis of osteomalacia. A rib or the iliac crest is the site at which biopsy is usually performed. To assess osteoid content and mineral appositional and bone formation rates, the bone biopsy specimen is processed without decalcification. This requires special equipment. In osteomalacia, bone is mineralized poorly and slowly, resulting in wide osteoid seams (>12 μm) and a large fraction of bone covered by unmineralized osteoid. States of high bone turnover (increased bone formation and resorption), such as hyperparathyroidism, can also cause wide osteoid seams and an increased osteoid surface, producing a superficial resemblance to osteomalacia. Therefore, the rate of bone turnover should be determined by labeling bone with tetracycline, which provides a fluorescent marker of the calcification front. When two doses of tetracycline are given at different times, the distance between the two labels divided by the time interval between the two doses equals the mineral appositional rate. The normal appositional rate is approximately 0.74 μm/d. Mineralization lag time—the time required for newly formed osteoid to be mineralized—can be calculated by dividing osteoid seam width by the appositional rate corrected by the linear extent of mineralization of the calcification front (a measure of the bone surface that is undergoing active mineralization as measured by tetracycline incorporation). Mineralization lag time is normally about 20 to 25 days. Bone formation rate is calculated as the product of the appositional rate times the linear extent of mineralization of the calcification front. Depressed appositional rate, increased mineralization lag time, and reduced bone formation rate clearly distinguish osteomalacia from high-turnover states such as hyperparathyroidism. Low-turnover states, as can be seen in various forms of osteoporosis, also have low appositional and bone formation rates, but these conditions are distinguished from osteomalacia by normal or reduced osteoid surface and volume.

C. Imaging studies   The radiologic features of rickets can be quite striking, especially in the young child. In growing bone, the radiolucent epiphyses are wide and flared, with irregular epiphyseal-metaphyseal junctions. Long bones may be bowed. The cortices of the long bones are often indistinct. Occasionally, evidence of secondary hyperparathyroidism—subperiosteal resorption in the phalanges and metacarpals and erosion of the
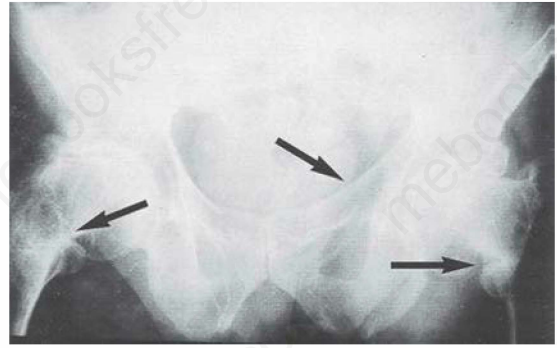
FIGURE 8–29   X-ray of the pelvis of an elderly woman with osteomalacia. Note marked bowing of both femoral necks, with pseudofractures of the medial aspect of the femoral necks and the superior aspect of the left pubic ramus (arrows). (Used with permission from Dr. Harry Genant.)

distal ends of the clavicles—is observed. Pseudofractures (also known as Looser zones or Milkman fractures) are an uncommon but nearly pathognomonic feature of rickets and osteomalacia (Figure 8–29). These radiolucent lines are most often found along the concave side of the femoral neck, pubic rami, ribs, clavicles, and lateral aspects of the scapulae. Pseudofractures may result from unhealed microfractures at points of stress or at the entry point of blood vessels into bone. They may progress to complete fractures that go unrecognized and thus lead to substantial deformity and disability. BMD is not a reliable indicator of osteomalacia, because BMD can be decreased in patients with vitamin D deficiency or increased in patients with chronic renal failure. In adults with normal renal function, radiologic evidence of a mineralization defect is often subtle and not readily distinguished from osteoporosis, with which it often coexists.

Treatment

The treatment of vitamin D deficiency is covered under that heading. The treatment of other diseases causing rickets or osteomalacia is found below under the specific disease headings.

NEPHROTIC SYNDROME

Nephrotic syndrome may lead to osteomalacia because of losses of vitamin D metabolites in the urine. Vitamin D metabolites are tightly bound to DBP, an α-globulin, and less tightly bound to albumin as described earlier. Patients with the nephrotic syndrome may lose large amounts of DBP and albumin in their urine and so deplete their vitamin D stores. Such patients can have very low levels of vitamin D metabolites in their serum, although the free concentrations are less affected. Thus, the measurement of total 25(OH)D or 1,25(OH)$_2$D may be misleading with respect to the severity of the vitamin D deficiency. Although bone disease has been recognized as a complication of the nephrotic syndrome, the prevalence of osteomalacia in this population is unknown. If vitamin D deficiency is suspected, treatment with vitamin D is

indicated with the proviso that normal total levels of 25(OH)D and $1,25(OH)_2D$ are not the goal. Serum and urine levels of calcium and phosphate, and plasma PTH levels are a more reliable guide to treatment.

HEPATIC OSTEODYSTROPHY

Hepatic osteodystrophy is the bone disease associated with liver disease. In the United States, patients with severe liver disease generally have osteoporosis, not osteomalacia. Both osteomalacia and osteoporosis are found in patients with liver disease in Great Britain, where vitamin D deficiency is more common. Although the liver is the site of the first step in the bioactivation of vitamin D—the conversion of vitamin D to 25(OH)D—this process is not tightly controlled, and until the liver disease is severe it is not rate limiting. Neither cholestatic nor parenchymal liver disease has much effect on 25(OH)D production until the late stages of liver failure. The low levels of 25(OH)D and $1,25(OH)_2D$ associated with liver disease can usually be attributed to the reduced production of DBP and albumin, poor nutrition, or malabsorption rather than a deficiency in the vitamin D 25-hydroxylase. As in patients with the nephrotic syndrome, the low total levels of the vitamin D metabolites may be misleading, as they reflect a reduction in DBP and albumin rather than a reduction in the free concentrations of these metabolites, which in several studies have been shown to be normal.

DRUG-INDUCED OSTEOMALACIA

Phenytoin and phenobarbital are anticonvulsants that induce drug-metabolizing enzymes in the liver that alter the hepatic metabolism of vitamin D and its metabolites. This action is not limited to anticonvulsants; the antituberculosis drug rifampin has been reported to do likewise. This effect may account for the lower circulating levels of 25(OH)D found in patients treated with such drugs. Levels of $1,25(OH)_2D$ are less affected. These therapies do not appear to lead to clinically significant bone disease except in subjects with other predisposing factors such as inadequate sunlight exposure (institutionalized patients) and poor nutrition. Children may be more vulnerable than adults. In animal studies, phenytoin has been noted to exert a direct inhibitory effect on bone mineralization, but the relevance of this observation to the human use of this drug is uncertain. The decrease in 25(OH)D levels in patients taking these drugs can be readily reversed with supplemental vitamin D administration.

HYPOPHOSPHATEMIC DISORDERS

Chronic hypophosphatemia may lead to rickets or osteomalacia independently of other predisposing abnormalities. The principal diseases in which hypophosphatemia is associated with osteomalacia or rickets, however, also include other abnormalities that can interfere with bone mineralization. Chronic phosphate depletion is caused by dietary deficiency (as in strict vegetarians), decreased intestinal absorption, or increased renal clearance (renal wasting). Acute hypophosphatemia can result from movement of phosphate into cells (eg, after infusion of insulin and glucose), but this condition is transient and does not result in bone disease.

Seventy to ninety percent of dietary phosphate is absorbed under normal conditions in the jejunum. This process is not tightly regulated, although $1,25(OH)_2D$ stimulates phosphate absorption, a factor that needs to be considered when calcitriol is being used to treat other conditions. Meat and dairy products are the principal dietary sources of phosphate. The incidence of osteomalacia in vegetarians who avoid all meat and dairy products is unknown, but when combined with low vitamin D production (decreased sunlight exposure) or intake as in East Asian immigrants in temperate countries such as England, osteomalacia is common. Intrinsic small bowel disease and small bowel surgery interfere with phosphate absorption and, if coupled with diarrhea or steatorrhea, can result in phosphate depletion. Antacids that contain aluminum hydroxide can bind phosphate and prevent its absorption. Patients who ingest large amounts of these antacids may become phosphate depleted. When this occurs in the setting of renal failure, reductions in serum phosphate plus aluminum intoxication in these individuals, who also have reduced $1,25(OH)_2D$ production, can produce profound osteomalacia. Fortunately this complication of these drugs is now well recognized, and use of such antacids as phosphate binders has declined. Eighty-five to ninety percent of the phosphate filtered by the glomerulus is reabsorbed, primarily in the proximal tubule. This process is regulated by PTH which reduces renal tubular phosphate reabsorption. It is also probably regulated by various vitamin D metabolites that appear to increase renal tubular phosphate reabsorption, and importantly by FGF23, which blocks the insertion of sodium phosphate cotransporters into the luminal membrane of renal proximal tubular cells, thus limiting renal phosphate reabsorption. Many diseases that affect renal handling of phosphate are associated with osteomalacia—especially those also associated with abnormalities in vitamin D metabolism and in FGF23 production and/or clearance, as discussed later (see also Chapter 21).

Treatment of phosphate deficiency is generally geared to correction of the primary problem. Oral preparations of phosphate (and the amounts required to provide 1 g of elemental phosphorus) include Fleet Phospho-soda (6.12 mL) and Neutra-Phos (300 mL). These preparations are usually given in amounts that provide 1 to 3 g of phosphorus daily in divided doses, although diarrhea may limit the dose. Careful attention to both serum calcium and serum phosphate concentrations is required to avoid hypocalcemia, ectopic calcification, or secondary hyperparathyroidism.

X-Linked and Autosomal Dominant Hypophosphatemia

X-linked hypophosphatemia (XLH), formerly called vitamin D-resistant rickets, is characterized by renal phosphate wasting, hypophosphatemia, and decreased $1,25(OH)_2D$ production relative to the degree of hypophosphatemia. Clinical presentation is variable, but children often present with florid rickets. This X-linked

dominant disorder generally affects males more severely than females. A similar but genetically distinct syndrome, autosomal dominant hypophosphatemic rickets (ADHR), is less common, affects females to the same extent as males, and may present later in life (second to fourth decades). The primary cause of hypophosphatemia in both conditions is excess FGF23. Circulating FGF23 is responsible for both the reduction in phosphate reabsorption in the renal proximal tubule and the blunted response of the proximal tubule to the ensuing hypophosphatemia with respect to $1,25(OH)_2D$ production. The gene responsible for XLH is *PHEX*. It shares structural homology with a number of endopeptidase genes and is located on the X chromosome. As is implicit in this ambiguous name, the role of PHEX in renal phosphate transport is not clear. The prevailing hypothesis is that PHEX indirectly leads to the cleavage and subsequent inactivation of FGF23, thus relieving the inhibition of phosphate reabsorption and $1,25(OH)_2D$ production.

ADHR is due to a mutation in FGF23 that renders it resistant to cleavage. Recent evidence suggests that the defects in XLH and ADHR may not be restricted to the proximal renal tubule but may affect osteoblast function as well, a cell in which *PHEX* is expressed.

Treatment of XLH and ADHR generally requires a combination of phosphate (1-4 g/d in divided doses) combined with calcitriol (1-3 μg/d) as tolerated. Vitamin D is less effective than calcitriol and should not be used for this condition. Recent studies with a neutralizing antibody to FGF23 in patients with XLH show promise and may help reduce the doses of calcitriol and phosphate required to treat this condition and ultimately achieve better control of the biochemical parameters. Appropriate therapy heals the rachitic lesions and increases growth velocity. A frequent complication of these diseases, however, is the development of hyperparathyroidism, which may become autonomous and require parathyroidectomy. It is unclear whether the hyperparathyroidism is a complication of therapy or intrinsic to the disease process itself. Treatment with phosphate and calcitriol may also lead to ectopic calcification, including nephrocalcinosis. Thus, these patients need to be followed closely.

Tumor-Induced Osteomalacia

The association of osteomalacia and hypophosphatemia with tumors, primarily of mesenchymal origin, has been recognized for half a century (see Chapter 21). The cause of this syndrome has until recently remained obscure. Removal of the tumors, which are often small and difficult to find, cures the disease. Although most implicated tumors are mesenchymal, including fibromas and osteoblastomas, other tumors, including breast, prostate, and lung carcinomas; multiple myeloma; and chronic lymphocytic leukemia have been associated with this syndrome. The patient generally presents with bone pain, muscle weakness, and osteomalacia. Symptoms may occur for years before the diagnosis is made. Renal phosphate wasting, hypophosphatemia, and normal serum calcium and 25(OH)D levels but inappropriately low $1,25(OH)_2D$ levels characterize the disease. Thus, it resembles XLH and ADHR. Tumor extracts contain molecules that inhibit renal phosphate transport and $1,25(OH)_2D$ production, suggesting that a phosphate-regulating substance or **phosphatonin** is involved. Thus, it was of great interest to observe that many of these tumors overproduce FGF23 and that production by the tumor of FGF23 can explain the metabolic abnormalities associated with the hypophosphatemic syndrome (see Figure 8–10). Other tumor products such as MEPE (matrix extracellular phosphoglycoprotein) and sFRP4 (secreted frizzled related protein 4) have also been implicated. The best treatment is removal of the tumor, but this is not always possible. Phosphate and calcitriol have been the mainstays of treatment but with limited success. Patients should be dosed to tolerance as in XLH and ADHR.

Fibrous Dysplasia

Fibrous dysplasia results from activating mutations in the *GNAS* gene, which encodes Gs alpha. Fibrous dysplasia of bone is a component of the McCune-Albright syndrome. The bony lesions in this disorder overproduce FGF23 which like tumor-induced osteomalacia, described earlier, leads to hypophosphatemia, osteomalacia, and inappropriately normal $1,25(OH)_2D$. These lesions can be mono-ostotic or polyostotic, with the extent of bone involved correlating with the degree of FGF23 hypersecretion, and hence the hypophosphatemia observed.

De Toni-Debré-Fanconi Syndrome and Hereditary Hypophosphatemic Rickets with Hypercalciuria

The De Toni-Debré-Fanconi syndrome includes a heterogeneous group of disorders affecting the proximal tubule and leading to phosphaturia and hypophosphatemia, aminoaciduria, glycosuria, bicarbonaturia, and proximal renal tubular acidosis. Not all features must be present to make the diagnosis. Damage to the renal proximal tubule secondary to genetic or environmental causes is the underlying cause. Drugs that cause the renal Fanconi syndrome include antiretrovirals such as tenofovir and adefovir (see Chapter 25), the alkylating agent ifosfamide, cisplatin, valproic acid, and the antibiotics gentamicin and tetracycline. Genetic causes of an autosomal recessive Fanconi syndrome include a duplication of *SLA34A1,* which encodes NaPiIIa, but how this leads to a generalized proximal tubular dysfunction is not clear. This syndrome may be divided into two types depending on whether vitamin D metabolism is also abnormal. In the more common type I, $1,25(OH)_2D$ production is reduced relative to the degree of hypophosphatemia, similar to that seen in XLH. In type II, $1,25(OH)_2D$ production is appropriately elevated in response to the hypophosphatemia, and this leads to hypercalciuria, similar to that seen in hereditary hypophosphatemic rickets with hypercalciuria (HHRH). The reason that $1,25(OH)_2D$ production is reduced in type I, despite the hypophosphatemia, is unclear but may suggest damage to the mitochondria where the 1-hydroxylase is located.

As noted earlier, the type II Fanconi syndrome shares the principal features of HHRH. This latter rare disorder results from mutations in *SLC34A3,* the gene that encodes the type IIc sodium

phosphate cotransporter (NaPiIIc), leading to loss of function of this transporter. Mutations in NaPiIIa, the other major renal sodium phosphate cotransporter, produces a similar picture. The key difference between HHRH and ADHR or XLH is that FGF23 is not elevated—so 1-hydroxylase activity is not suppressed. Thus, the hypophosphatemia leads to increased $1,25(OH)_2D$ production, increased intestinal calcium absorption, and hypercalciuria. HHRH is heterogeneous in clinical presentation, but in severely affected individuals, it begins in childhood with bone pain and skeletal deformities. The syndrome is characterized by renal phosphate wasting and hypophosphatemia, hypercalciuria and normal serum calcium, and elevated $1,25(OH)_2D$ levels.

The osteomalacia associated with these renal phosphate-wasting syndromes is probably the result of the hypophosphatemia, with contributions from the acidosis and decreased $1,25(OH)_2D$ levels in the type I syndrome.

The ideal treatment is correction of the underlying defect, which may or may not be identified and/or reversible. Otherwise, treatment includes phosphate supplementation in all cases, correction of the acidosis, and $1,25(OH)_2D$ replacement in the type I syndrome, but not in HHRH as this will aggravate the hypercalciuria.

CALCIUM DEFICIENCY

Calcium deficiency may contribute to the mineralization defect that complicates gastrointestinal disease and proximal tubular disorders, but it is less well established as a cause of osteomalacia than is vitamin D or phosphate deficiency. In carefully performed studies of children who ingested a very low-calcium diet in Africa, there was clinical, biochemical, and histologic evidence of osteomalacia. The serum phosphate, 25(OH)D and $1,25(OH)_2D$ levels were normal; serum alkaline phosphatase levels were elevated; and serum and urine calcium levels were low. Because the efficiency of intestinal absorption of calcium decreases with age, the daily requirement for calcium increases from approximately 1000 mg in young adults to 1200 mg in the elderly. Calcium deficiency can result not only from inadequate dietary intake but also from excessive fecal and urinary losses. Except in cases in which a renal leak of calcium plays an important role in the etiology of calcium deficiency (certain forms of idiopathic hypercalciuria or following glucocorticoid therapy for inflammatory diseases), urinary calcium excretion provides a useful means to determine the appropriate level of oral calcium replacement. Because of its low cost and high percentage of elemental calcium, calcium carbonate is the treatment of choice, although calcium citrate may be better absorbed on an empty stomach and is less constipating.

PRIMARY DISORDERS OF THE BONE MATRIX

Several diseases lead to abnormalities in bone mineralization because of intrinsic defects in the matrix or in the osteoblast producing that matrix. With the exception of osteogenesis imperfecta, these tend to be rare. Examples are listed later.

OSTEOGENESIS IMPERFECTA

The classic forms of osteogenesis imperfecta (OI) are heritable disorders of connective tissue due to qualitative or quantitative abnormalities in type I collagen, the most abundant collagen in bone. The disease is typically transmitted in an autosomal dominant mode, although autosomal recessive inheritance has also been described in the forms of OI due to mutations in genes encoding noncollagenous proteins. The clinical phenotype of OI is highly variable, depending on the location and type of mutation observed. Although skeletal deformity and fracture are the hallmarks of this disease, other tissues are often affected, including the teeth, skin, ligaments, and scleras. Routine biochemical studies of bone and mineral metabolism are generally normal; however, elevated serum alkaline phosphatase and increased excretion of biochemical markers of bone turnover and calcium are often found. Bone histology shows abundant disorganized, poorly mineralized matrix but sometimes with a high rate of bone turnover as reflected by increased tetracycline labeling.

Treatment is supportive and includes orthopedic, rehabilitative, and dental intervention as appropriate. Bisphosphonates have recently been shown to be of value in children to prevent fractures and reduce bone pain and skeletal morbidity. Bisphosphonates have also been tested used in adults, often in the setting of fragility fractures, as well as teriparatide as a therapy to rebuild bone microarchitecture and mass.

Hypophosphatasia

Hypophosphatasia, transmitted in an autosomal recessive or dominant pattern, has considerable variability in clinical expression ranging from a severe form of rickets in children to a predisposition to fractures in adults. The biochemical hallmarks are low serum and tissue levels of alkaline phosphatase (liver-bone-kidney but not intestine-placenta) and increased urinary levels of phosphoethanolamine. Serum calcium and phosphate levels may be high, and many patients have hypercalciuria. Levels of vitamin D metabolites and PTH are generally normal. In some families, a mutation in the *tissue nonspecific alkaline phosphatase* gene has been identified. Radiologic examination shows osteopenia, with the frequent occurrence of stress fractures and chondrocalcinosis. Bone biopsy specimens show osteomalacia. Skeletal alkaline phosphatase cleaves pyrophosphate, an inhibitor of bone mineralization. Thus, patients deficient in alkaline phosphatase may be unable to hydrolyze this inhibitor and so develop a mineralization defect. Current therapy in severe cases involves the administration of asfotase alfa, a recombinant form of alkaline phosphatase.

Fibrogenesis Imperfecta Ossium

Fibrogenesis imperfecta ossium is a rare, painful disorder that affects middle-aged subjects in what appears to be a sporadic fashion. Serum alkaline phosphatase activity is increased. The bones have a dense, amorphous mottled appearance radiologically and a disorganized arrangement of collagen with decreased birefringence

histologically when viewed with polarized light microscopy. Presumably, the disorganized collagen matrix retards normal bone mineralization. There is no specific therapy.

INHIBITORS OF MINERALIZATION

Several drugs are known to cause osteomalacia or rickets by inhibiting mineralization. However, the mechanisms by which this occurs are not fully understood.

Aluminum

Aluminum-induced bone disease is found primarily in patients with renal failure on chronic hemodialysis or in patients being treated with total parenteral nutrition (TPN). With the recognition of this complication and the switch to deionized water for dialysis, elimination of aluminum-containing antacids, and the use of purified amino acids rather than casein hydrolysates (which contained large amounts of aluminum) for TPN, the incidence and prevalence of this complication have been markedly reduced. The problem, however, still exists. As described under renal osteodystrophy, the bone disease is osteomalacia and can be severe and painful. Aluminum deposits in the mineralization front and appears to inhibit the mineralization process. Removal of the aluminum with deferoxamine chelation therapy is the treatment of choice.

Fluoride

Fluoride is a potent stimulator of bone formation. If administered at high doses with inadequate calcium supplementation, the bone formed as a result is poorly mineralized. The mechanisms underlying the effects of fluoride on osteoblast function and bone mineralization remain unclear.

PAGET DISEASE OF BONE (OSTEITIS DEFORMANS)

Paget disease is a focal disorder of bone remodeling that leads to greatly accelerated rates of bone turnover, disruption of the normal architecture of bone, and sometimes to gross deformities of bone. As a focal disorder it is not, strictly speaking, a metabolic bone disease. Paget disease is prevalent in northern Europe, particularly in England and Germany. It is also common in the United States but is unusual in Africa and Asia.

Etiology

It has long been thought that Paget disease, with its late onset and spotty involvement of the skeleton, might be due to a chronic slow virus infection of bone. Inclusion bodies that resemble paramyxovirus inclusions have been identified in pagetic osteoclasts, and the presence of measles virus transcripts has been detected by molecular techniques. However, considerably more work would be required to prove the infectious etiology of Paget disease. There are also familial clusters of the disease, with up to 20% of patients in some studies having afflicted first-degree relatives so several pathogenetic mechanisms are likely (discussed later).

Pathology

At the microscopic level, the disorder is characterized by highly vascular and cellular bone, consistent with its high metabolic activity. The osteoclasts are sometimes huge and bizarre, with up to 100 nuclei per cell. Because pagetic osteoclasts initiate the bone remodeling cycle in a chaotic fashion, the end result of remodeling is a mosaic pattern of lamellar bone. Paget disease (and other high turnover states) can also produce woven bone—bone that is laid down rapidly and in a disorganized fashion, without the normal lamellar architecture.

Pathogenesis

Abnormal osteoclastic bone resorption is the probable initiating event in Paget disease. Not only are the osteoclasts abnormal histologically and prone to unruly behavior, but some forms of the disease are marked by an early resorptive phase in which pure osteolysis occurs without an osteoblastic response. Additionally, Paget disease responds dramatically to inhibitors of osteoclastic bone resorption.

The rate of bone resorption is often increased by as much as 10- to 20-fold, and this is reflected in biochemical indices of bone resorption. Over the skeleton as a whole, osteoblastic new bone formation responds appropriately to this challenge. Even though local disparities in remodeling may result in areas with the radiographic appearance of osteolysis or dense new bone, there is a close relationship between biochemical markers of bone resorption (eg, N-telopeptide, C-telopeptide, deoxypyridinoline) and formation (eg, alkaline phosphatase, osteocalcin). Because this tight coupling is maintained in Paget disease, in the face of enormously increased skeletal turnover rates, systemic mineral homeostasis is usually unperturbed.

Genetic Forms

Approximately 20% of patients with Paget disease have a positive family history. This interesting observation plus the uncertain viral etiology has led several groups to determine the genetic loci responsible for the disease. These studies plus work done to unravel the molecular defects in patients with rare but clear disorders of accelerated and unregulated bone remodeling have shed light on molecular pathways that may be important in sporadic Paget disease.

Mutations in RANK and OPG are important in the pathogenesis of high bone remodeling states. Whyte and colleagues investigated two unrelated Navajo patients with autosomal recessive juvenile Paget disease in whom they found approximately 100-kb deletions of the gene encoding TNFRSF11B or OPG on chromosome 8q24.2. In two other clinical disorders of high bone remodeling designated as familial expansile osteolysis and expansile skeletal hyperphosphatasia, tandem duplications are present within the signal peptide of RANK, leading to a

gain-of-function in the activity of RANK. This is predicted to enhance osteoclastogenesis.

Familial Paget disease is not, however, generally explained by abnormalities in *RANK* or *OPG* genes. In contrast, several groups have uncovered mutations in SQSTM1 or sequestosome 1 (also known as p62) in multiple kindreds with Paget disease. SQSTM1 is a ubiquitin-binding protein involved in signaling pathways for IL-1, TNF, and RANKL—all molecules important in the production and function of osteoclasts. A variety of different SQSTM1 mutations are thought to explain approximately 30% of cases of familial Paget disease. The identification of additional candidate genes and their roles in bone metabolism are areas of active investigation.

Clinical Features

A. Symptoms and signs   Typical Paget disease may affect any bone, but the most common sites are the sacrum and spine (50% of patients), femur (46%), skull (28%), and pelvis (22%). The clinical features of Paget disease are pain, fractures, deformity, and manifestations of the neurologic, rheumatologic, or metabolic complications of the disease. However, at least two-thirds of patients are asymptomatic. Thus, Paget disease is often discovered as an incidental radiologic finding or during the investigation of an elevated alkaline phosphatase level. On physical examination, enlargement of the skull, frontal bossing, or deafness may be evident. Involvement of the weight-bearing long bones of the lower extremity often results in bowing. The femur and tibia bow anteriorly and laterally, but the fibula is almost never affected. Cutaneous erythema and warmth, as well as bone tenderness, may be evident over affected areas of the skeleton, reflecting greatly increased blood flow through pagetic bone. The findings of pain, warmth, and erythema led to the appellation osteitis deformans, although Paget disease is not truly an inflammatory disorder. The most common fractures in Paget disease are vertebral crush fractures and incomplete *fissure* fractures through the cortex, usually on the convex surface of the tibia or femur. Affected bones may fracture completely; when they do, healing is usually rapid and complete—the increased metabolic activity of pagetic bone seems to favor fracture healing.

B. Laboratory findings   The serum alkaline phosphatase activity is usually increased, sometimes to very high levels. Levels of biochemical markers of bone turnover are often high, but it is not clear that their determination offers any special advantage over that of serum alkaline phosphatase activity. The serum calcium and phosphate concentrations and the urinary calcium excretion are normal, although if a patient sustains a fracture and becomes immobilized, hypercalciuria and hypercalcemia may occur.

C. Imaging studies   The early stages of Paget disease are often osteolytic. Examples are erosion of the temporal bone of the skull, osteoporosis circumscripta, and pagetic lesions in the extremities, which begin in the metaphysis and migrate down

FIGURE 8–30   Lytic Paget disease in the tibia before (**left**) and after (**right**) immobilization in a cast. The lytic area has a flame- or V-shaped leading edge (**left**). (Reproduced with permission from Strewler GJ. Paget disease of bone. *West J Med*. 1984 May;140(5): 763-768.)

the shaft as a V-shaped resorptive front (Figure 8–30). Over years or even decades, the typical mixed picture of late Paget disease evolves. Trabeculae are thickened and coarse. The bone may be enlarged or bowed. In the pelvis, the iliopectineal line or pelvic brim is often thickened (Figure 8–31). In the spine, osteoblastic lesions of the vertebral bodies may present a *picture-frame* appearance or a homogeneously increased density, the *ivory vertebra*. Associated osteoarthritis may present with narrowing of the joint space (see Figure 8–31). Osteosarcoma may present with cortical destruction or a soft tissue mass (see Figure 8–31). Radionuclide bone scanning with technetium-labeled bisphosphonates or other bone-seeking agents is uniformly positive in active Paget disease and is useful for surveying the skeleton when a focus of Paget disease has been found radiographically (Figure 8–32) or the disease is suspected.

Complications

Complications of Paget disease may be neurologic, otologic, rheumatologic, neoplastic, or cardiac (Table 8–15).

A. Neurologic/Otologic   The brain, spinal cord, and peripheral nerves are all at risk.

Sensorineural deafness occurs in up to 50% of patients in whom the skull is involved, and compression of the other cranial
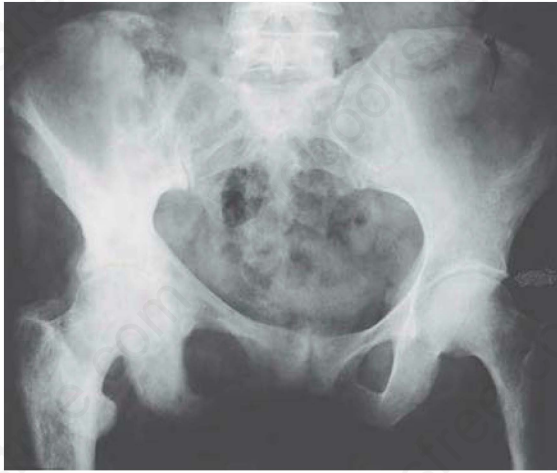
FIGURE 8–31   Paget disease of the right femur and pelvis. The right femur displays cortical thickening and coarse trabeculation. The right ischium is enlarged, with sclerosis of ischial and pubic rami and the right ilium. Two complications of Paget disease are present. There is concentric bilateral narrowing of the hip joint space, signifying osteoarthritis. The destructive lesion interrupting the cortex of the right ilium is an osteosarcoma. (Reproduced with permission from Strewler GJ. Paget disease of bone. *West J Med*. 1984 May;140(5): 763-768.)

nerves can also occur. At the base of the skull, Paget disease can produce platybasia and basilar impression of the brain stem, with symptoms of brain stem compression, obstructive hydrocephalus, or vertebrobasilar insufficiency. Spinal stenosis is common in vertebral Paget disease, in part because the pagetic vertebra can be

FIGURE 8–32   Bone scan of a patient with Paget disease of the pelvis, right femur, and acetabulum. Note localization of bone-seeking isotope (^{99m}Tc-labeled bisphosphonate) in these areas.

TABLE 8–15   Complications of Paget disease.

Rheumatologic Osteoarthritis Gout Calcific periarthritis
Neurologic/Otologic Basilar impression Hydrocephalus (rare) Cranial nerve dysfunction (especially deafness) Spinal cord and root compression Peripheral nerve entrapment (carpal and tarsal tunnel syndromes)
Metabolic Immobilization hypercalciuria-hypercalcemia Urolithiasis
Neoplastic Sarcoma Giant cell tumor

enlarged, and may extend posteriorly when collapse occurs, but spinal stenosis responds well to medical treatment of the disease.

B. Rheumatologic   Osteoarthritis is common in Paget disease. It may be an unrelated finding in elderly patients with the disorder, or it may result directly from pagetic deformities and their effects on wear and tear in the joints. Arthritis presents a conundrum to the clinician attempting to relieve pain, as it may be difficult to determine whether the pain originates in pagetic bone or in the nearby joint. An association of osteitis deformans and gout was first noted by James Paget himself, and asymptomatic hyperuricemia is also common.

C. Neoplastic   The most dreaded complication of Paget disease is development of a sarcoma. The tumor arises in pagetic bone, typically in individuals with polyostotic involvement, and may present with soft tissue swelling, increased pain, or a rapidly increasing alkaline phosphatase. Osteosarcoma, chondrosarcoma, and giant cell tumors all occur in Paget disease, with a combined incidence of about 1%. Because osteosarcoma is otherwise uncommon in the elderly, fully 30% of elderly patients with osteosarcoma have underlying Paget disease.

D. Cardiac   High-output congestive heart failure occurs rarely and is due to markedly increased blood flow to bone, usually in patients with more than 50% involvement of the skeleton.

Treatment

In a patient known to have Paget disease, bone pain that is unresponsive to nonsteroidal anti-inflammatory agents deserves a trial of specific therapy. As noted earlier, it may not be easy to differentiate pagetic bone pain from that due to osteoarthritis. The other indications for treatment of Paget disease are controversial. Treatment has been advocated for neurologic compression syndromes, as preparation for surgery, and to prevent deformities. Neurologic deficits often respond to medical treatment, and a trial of

treatment is often warranted. Pretreatment for 2 to 3 months before orthopedic surgery prevents excessive bleeding and postoperative hypercalcemia, but satisfactory bone healing usually occurs without medical treatment. Paget disease is sometimes treated in the hope of arresting the progress of deformities (eg, bowing of the extremities and resultant osteoarthritis), but it is not certain whether medical treatment can achieve this aim or whether it will arrest the progression of deafness in patients with skull involvement.

Two classes of agents are used in the treatment of Paget disease: bisphosphonates and CT. Both inhibit osteoclastic bone resorption.

A. Bisphosphonates The oral bisphosphonates (alendronate and risedronate) are sometimes used in the treatment of Paget disease. Alendronate is administered at a dose of 40 mg daily for 6 months. On average, alkaline phosphatase activity is suppressed by about 80% using such treatment. Biochemical remissions are often prolonged for more than 1 year, after the drug is stopped. The main side-effect is gastrointestinal upset, requiring discontinuation of therapy in about 6% of patients. Risedronate, given at a dosage of 30 mg/d for 2 months, is also highly efficacious in lowering and even normalizing alkaline phosphatase activity and reducing bone pain. For patients with gastrointestinal intolerance to oral bisphosphonates or more severe bony involvement, it may be more convenient or appropriate to administer intravenous bisphosphonates. If the disease is highly active and surgery is planned in the immediate future, intravenous therapy is generally preferred. Pamidronate, an aminobisphosphonate closely related to alendronate, has been used for this purpose. Intravenous infusions of pamidronate (60 or 90 mg) produce a high remission rate and a durable response. Preferably, intravenous zoledronic acid (5 mg) is the first-line treatment choice for efficacy because of both the strength and duration of alkaline phosphatase suppression, in patients with Paget disease. The principal side-effect observed with intravenous aminobisphosphonate administration is an acute-phase response, including fever and myalgias, that occurs in about 10% of patients and may last for several days after a dose. A very small percentage of patients receiving multiple and typically high doses of intravenous bisphosphonates can develop osteonecrosis of the jaw, although the majority of patients given intravenous zoledronic acid do not require retreatment.

B. Calcitonin Salmon CT is administered initially at a dosage of 50 to 100 IU daily until symptoms are improved; thereafter, many patients can be maintained on 50 units three times a week. Improvement in pain is usually evident within 2 to 6 weeks. On average, the alkaline phosphatase falls by 50% within 3 to 6 months. Many patients have a sustained response to treatment extending over years, and biochemical parameters are often suppressed for 6 months to 1 year after treatment is discontinued. Up to 20% of patients receiving chronic CT treatment develop late resistance to CT, which may be antibody-mediated. Data on the use of nasal spray CT's efficacy in Paget disease are limited.

BONE DISEASE IN CHRONIC KIDNEY DISEASE

Pathogenesis

Metabolism of $25(OH)D$ to $1,25(OH)_2D$ and $24,25(OH)_2D$ in the kidney is tightly regulated. CKD results in reduced circulating levels of both of these metabolites. Likely even earlier than the fall in $1,25(OH)_2D$ production that occurs in CKD is the rise in FGF23, perhaps due to phosphate retention by the diseased kidney. FGF23 rises progressively and dramatically with more advanced states of CKD. This increase contributes to the early fall in renal $1,25(OH)_2D$ production. High FGF23 levels may contribute to some of the cardiovascular complications of CKD like heart failure. Reduced $1,25(OH)_2D$ levels in progressive CKD cause intestinal calcium absorption to fall, and bone resorption appears to become less sensitive to PTH—a result that leads to hypocalcemia. Decreased phosphate excretion results in hyperphosphatemia, which amplifies the fall in serum calcium and independently increases PTH secretion. The fall in serum calcium combined with the low levels of $1,25(OH)_2D$, which is an inhibitor of PTH secretion, and the hyperphosphatemia result in hyperparathyroidism (Figure 8–33). The parathyroid glands, under chronic proliferative stimulation in uremic individuals, can develop clonal expansion of cells that harbor mutations in known parathyroid growth-promoting genes. This can result in nodular hyperplasia. In addition, the expression of both the VDR and CaSR is often depressed in tissues from individuals with chronic secondary hyperparathyroidism. Both calcium, likely acting via CaSRs, and $1,25(OH)_2D$, acting via VDRs, exert inhibitory effects on parathyroid cell growth.

The net effect of deficient $1,25(OH)_2D$ and $24,25(OH)_2D$ and excess PTH and FGF23 on bone in CKD is complex. Bone
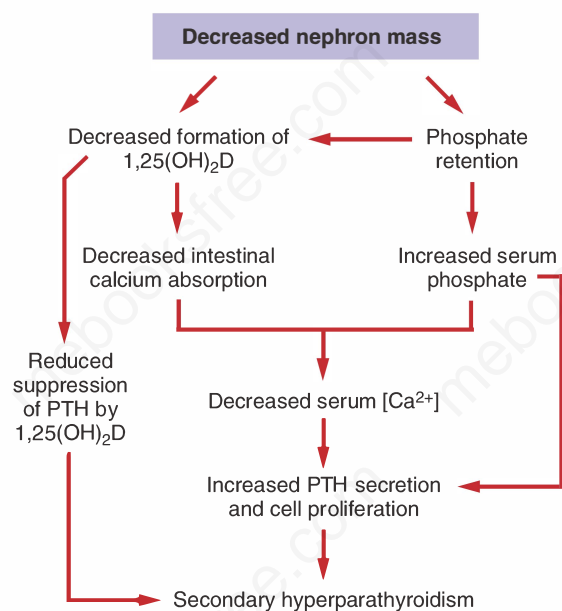
FIGURE 8–33 Pathogenesis of renal secondary hyperparathyroidism.

may show osteitis fibrosa cystica (reflecting excessive PTH), osteomalacia (in part reflecting decreased vitamin D metabolites and high FGF23 which have complex effects on bone mineralization), or a combination of the two. One particularly debilitating skeletal complication is low turnover or adynamic bone disease, now found in an increasing percentage of patients on hemodialysis. In the past but rarely now, some patients had increased aluminum content in their bones, particularly in the region where mineralization is occurring (calcification front), and this appeared to inhibit mineralization. More commonly, however, these patients do not have aluminum excess or increased osteoid but just have low bone turnover for reasons that are not entirely clear, although low levels of PTH have been implicated. Such patients are particularly prone to develop symptoms of bone pain, fractures, and muscle weakness, and are particularly sensitive to developing hypercalcemia when treated with calcitriol despite the relatively low levels of PTH. The combined picture of disordered bone and mineral metabolism and the pathologic vascular calcification in renal disease has been termed CKD-mineral bone disease (CKD-MBD).

Clinical Features

Most patients with advanced stages of CKD have osteitis fibrosa alone or in combination with osteomalacia. If not well controlled, these patients will have a low serum level of calcium and high serum levels of phosphorus, alkaline phosphatase, and PTH. A few patients develop severe secondary hyperparathyroidism in which the PTH level increases dramatically and stays elevated even with restoration of the serum calcium to normal. Such patients are prone to developing hypercalcemia with the usual replacement doses of calcium (sometimes referred to as tertiary hyperparathyroidism). As noted earlier another subset of patients with CKD present with relatively normal levels of PTH and alkaline phosphatase. Their serum calcium levels are often elevated after treatment with small doses of calcitriol. These patients generally have adynamic bone disease on bone biopsy.

Treatment

Patients with renal bone disease generally respond to calcitriol (0.5-1 µg/d) or one of the vitamin D analogs paricalcitol [19-nor-1-alpha, $25(OH)_2D_2$] or doxercalciferol ($1\alpha OHD_2$), calcium supplementation (1-2 g/d), and phosphate restriction. Vitamin D supplementation is also recommended to ensure adequate $25(OH)$D levels (30 ng/mL). To avoid aluminum intoxication, sevelamer (Renagel), lanthanum carbonate (Fosrenol), and other nonaluminum phosphate binders are used to reduce intestinal phosphate absorption. To minimize the calcium-phosphate product, absorbable calcium-containing salts are used only at low doses to maintain sufficient calcium intake and not to act as a phosphate binder. The goal is to maintain normal serum levels of calcium, phosphate, and alkaline phosphatase. This regimen treats osteitis fibrosa more effectively than osteomalacia.

To prevent the development of adynamic bone disease, intact PTH levels higher than the upper limit of normal in standard intact PTH assays are encouraged. The optimal intact PTH level

for preserving skeletal integrity in end-stage renal disease (ESRD) and in patients on dialysis is uncertain. Newer or so-called biointact or whole PTH assays give a better reflection of the levels of the full-length PTH(1-84) molecule in these settings (described earlier). Patients with aluminum intoxication are very sensitive to and respond poorly to calcitriol and calcium, with rapid onset of hypercalcemia and little improvement in their bone disease. These patients may respond to chelation therapy with deferoxamine, a drug also used for iron chelation.

Severe secondary hyperparathyroidism is difficult to manage. These patients are prone to the development of hypercalcemia with calcium supplementation. Intravenous doses of calcitriol with hemodialysis may preferentially inhibit PTH secretion with less effect on raising serum calcium than when given orally. Several analogs of calcitriol with less hypercalcemic potential than calcitriol itself are approved for the management of such patients.

A major advance in the treatment of uremic secondary hyperparathyroidism is the use of the calcimimetic cinacalcet. This agent directly inhibits PTH secretion by activating the CaSR on parathyroid cells. Cinacalcet lowers levels of PTH and presumably of bone turnover and has less of a tendency to elevate the calcium-phosphate product, unlike the vitamin D analogs, which tend to raise serum calcium and phosphate. Cinacalcet has contributed substantially to the medical management of uremic secondary hyperparathyroidism and has markedly reduced the need for subtotal parathyroidectomy to treat refractory secondary or tertiary hyperparathyroidism in patients with CKD. The major side-effects of cinacalcet, nausea and vomiting, limit its use in some patients.

HEREDITARY FORMS OF HYPERPHOSPHATEMIA

Tumoral Calcinosis

Mutations in genes leading to reduced FGF23 production, secretion, or action have recently been described which result in hyperphosphatemia and increased $1,25(OH)_2D$ production. FGF23 requires glycosylation for its export from the cell where it is produced. Although FGF23 utilizes FGF receptors in common with many other members of the FGF family (FGFR 1 and 3), it also requires a coreceptor **alpha klotho**, to activate FGF23-mediated signaling in target cells. Inactivating mutations in FGF23 itself, the enzyme that glycosylates FGF23 which is necessary for secretion, **GALNT3** (UDP-N-acetyl-α-D-galactosamine transferase), and alpha klotho all lead to hyperphosphatemia and increased $1,25(OH)_2D$ levels, although serum calcium levels are generally normal. Urinary phosphate reabsorption is increased. These findings are the opposite of those of familial hypophosphatemic disorders described earlier. As a result of the increased calcium-phosphate product, ectopic calcification occurs, generally around joints and soft tissues. Dentition may be affected in some individuals. Treatment involves the use of phosphate binders in an effort to reduce the serum phosphate levels and reduce the painful and sometimes disfiguring ectopic calcifications which can become very large.

REFERENCES

Parathyroid Hormone, Parathyroid Hormone–Related Peptide, and Their Receptors

Cheloha RW, Gellman SH, Vilardaga J-P, Gardella TJ. PTH receptor-1 signalling-mechanistic insights and therapeutic prospects. *Nat Rev Endocrinol*. 2015;11: 712. [PMID: 26303600]

D'Amour P. Acute and chronic regulation of circulating PTH: significance in health and in disease. *Clin Biochem*. 2012;45:964. [PMID: 22569597]

Martin TJ. Parathyroid hormone-related protein, its regulation of cartilage and bone development, and role in treating bone diseases. *Physiol Rev*. 2016;96: 831. [PMID: 27142453]

Calcitonin

Katafuchi T, Yasue H, Osaki T, Minamino N. Calcitonin receptor-stimulating peptide: its evolutionary and functional relationship with calcitonin/calcitonin-gene related peptide based on gene structure. *Peptides*. 2009;30:1753. [PMID: 19540291]

Naot D, Cornish J. The role of peptides and receptors of the calcitonin family in the regulation of bone metabolism. *Bone*. 2008;43:813. [PMID: 18687416]

Vitamin D

Bikle DD. Vitamin D metabolism, mechanism of action, and clinical applications. *Chem Biol*. 2014;21:319. [PMID: 2459992]

Christakos S, Dhawan P, Verstuyf A, Verlinden L, Carmeliet G. Vitamin D: metabolism, molecular mechanism of action, and pleiotropic effects. *Physiol Rev*. 2016;96:365. [PMID: 26681795]

Holick MF, Binkley NC, Bischoff-Ferrari HA, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96:1911. [PMID: 21646368]

IOM (Institute of Medicine) 2011 Dietary reference intakes for calcium and vitamin D. Committee to Review Dietary Reference Intakes for Calcium and Vitamin D. Washington, DC: National Academies Press, Institute of Medicine.

Pike JW, Meyer MB. Fundamentals of vitamin D hormone-regulated gene expression. *J Steroid Biochem Mol Biol*. 2014;144:5. [PMID: 24239506]

Medullary Thyroid Carcinoma

Bae YJ, Schaab M, Kratzsch J. Calcitonin as biomarker for the medullary thyroid carcinoma. *Rec Results Canc Res*. 2015;204:117. [PMID: 26494386]

Cote GJ, Grubbs EG, Hofmann MC. Thyroid C-cell biology and oncogenic transformation. *Rec Results Canc Res*. 2015;204:1. [PMID: 26494382]

Frank-Raue K, Raue F. Hereditary medullary thyroid cancer genotype-phenotype correlation. *Rec Results Canc Res*. 2015;204:139. [PMID: 26494387]

Hadoux J, Pacini F, Tuttle RM, Schlumberger M. Management of advanced medullary thyroid cancer. *Lancet Diabetes Endocrinol*. 2016;4:64. [PMID: 26608066]

Lairmore TC, Diesen D, Goldfarb M, et al. American Association of Clinical Endocrinologists and American College of Endocrinology disease state clinical review: timing of multiple endocrine neoplasia thyroidectomy and extent of central neck lymphadenectomy. *Endocrine Pract*. 2015;21:839. [PMID: 26172129]

Romei C, Ciampi R, Elisei R. A comprehensive overview of the role of the *RET* proto-oncogene in thyroid carcinoma. *Nat Rev Endocrinol*. 2016;12:192. [PMID: 26868437]

Wells Jr SA, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567. [PMID: 25810047]

Hypercalcemic Disorders

Agarwal SK, Mateo CM, Marx SJ. Rare germline mutations in cyclin-dependent kinase inhibitor genes in multiple endocrine neoplasia type 1 and related states. *J Clin Endocrinol Metab*. 2009;94:1826. [PMID: 19141585]

Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3561. [PMID: 25162665]

Bilezikian JP, Cusano NE, Khan AA, et al. Primary hyperparathyroidism. *Nat Rev Dis Primers*. 2016;2:16033. [PMID: 27194212]

Donovan PJ, Achong N, Griffin K, et al. PTHrP-mediated hypercalcemia: causes and survival in 138 patients. *J Clin Endocrinol Metab*. 2015;100:2024. [PMID: 25719931]

Donovan PJ, Sundac L, Pretorius CJ, d'Emden MC, McLeod DSA. Calcitriol-mediated hypercalcemia: causes and course in 101 patients. *J Clin Endocrinol Metab*. 2013;98:4023. [PMID: 23979953]

Duan K, Hernandez KG, Mete O. Clinicopathological correlates of hyperparathyroidism. *J Clin Pathol*. 2015;68:771. [PMID: 26163537]

Gastanaga VM, Schwartzberg LS, Jain RK, et al. Prevalence of hypercalcemia among cancer patients in the United States. *Cancer Medicine*. 2016;5:2091. [PMID: 26520619]

Griebeler ML, Kearns AE, Ryu E, et al. Thiazide-associated hypercalcemia: incidence and association with primary hyperparathyroidism over two decades. *J Clin Endocrinol Metab*. 2016;101:1166. [PMID: 26751196]

Hu MI, Glezerman IG, Leboulleux S, et al. Denosumab for the treatment of hypercalcemia of malignancy. *J Clin Endocrinol Metab*. 2014;99:3144. [PMID: 24915117]

Li Y, Simonds WF. Endocrine neoplasms in familial syndromes of hyperparathyroidism. *Endocr Rel Cancer*. 2016;23:R229. [PMID: 27207564]

Maier JD, Levine SN. Hypercalcemia in the intensive care unit: a review of pathophysiology, diagnosis, and modern therapy. *J Int Care Med*. 2016;30:235. [PMID: 2495117]

Marcocci C, Bollerslev J, Khan AA, Shoback DM. Medical management of primary hyperparathyroidism: proceedings of the Fourth International Workshop on the Management of Asymptomatic Primary Hyperparathyroidism. *J Clin Endocrinol Metab*. 2014;99:3607. [PMID: 25162668]

Marcocci C, Chanson P, Shoback D, et al. Cinacalcet reduces serum calcium concentrations in patients with intractable primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2009;94:2766. [PMID: 19470620]

Mirrakhimov AE. Hypercalcemia of malignancy: an update on pathogenesis and management. *N Am J Med Sci*. 2015;7:483. [PMID: 26713296]

Nesbit MA, Hannan FM, Howles SA, et al. Mutations affecting G-protein subunit alpha-11 in hypercalcemia and hypocalcemia. *N Engl J Med*. 2013;368:2476. [PMID: 238025016]

Nesbit MA, Hannan FM, Howles SA, et al. Mutations in AP2S1 cause familial hypercalcemia type 3. *Nat Genet*. 2013;45:93. [PMID: 23222959]

Peacock M, Bolognese MA, Borofsky M, et al. Cinacalcet treatment of primary hyperparathyroidism: biochemical and bone densitometric outcomes in a five-year study. *J Clin Endocrinol Metab*. 2009;94:4860. [PMID: 19837909]

Rubin MR, Bilezikian JP, McMahon DJ, et al. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years. *J Clin Endocrinol Metab*. 2008;93:3462. [PMID: 18544625]

Silverberg SJ, Clarke BL, Peacock M, et al. Current issues in the presentation of asymptomatic primary hyperparathyroidism: proceedings of the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3580. [PMID: 25162666]

Sternlicht H, Glezerman IG. Hypercalcemia of malignancy and new treatment options. *Ther Clin Risk Management*. 2015;11:1779. [PMID: 26675713]

Thakker RV. Genetics of parathyroid tumours. *J Intern Med*. 2016;280(6):574-583. [PMID: 27306766]

Udelsman R, Akerstrom G, Biagini C, et al. The surgical management of asymptomatic primary hyperparathyroidism: proceedings of the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3595. [PMID: 25162669]

Hypocalcemic Disorders

Anderson MS, Su MA. AIRE expands: new roles in immune tolerance and beyond. *Nat Rev Immunol*. 2016;16:247. [PMID: 26972725]

Brandi ML, Bilezikian JP, Shoback D, et al. Management of hypoparathyroidism: summary statement and guidelines. *J Clin Endocrinol Metab*. 2016;101:2273. [PMID: 26943719]

Clarke BL, Brown EM, Collins MT, et al. Epidemiology and diagnosis of hypoparathyroidism. *J Clin Endocrinol Metab*. 2016;101:2284. [PMID: 26943720]

Elli FM, Barbieri AM, Bordogna P, et al. Screening for GNAS genetic and epigenetic alterations in progressive osseous heteroplasia: first Italian series. *Bone*. 2013;56:276. [PMID: 23796510]

Hannan FM, Thakker RV. Investigating hypocalcemia. *BMJ.* 2013;346:f2213. [PMID: 23661111]

Kisand K, Lilic D, Casanova JL, et al. Mucocutaneous candidiasis and autoimmunity against cytokines in APECED and thymoma patients: clinical and pathogenetic implications. *Eur J Immunol.* 2011;41:1517. [PMID: 21574164]

Lemos MC, Thakker RV. GNAS mutations in pseudohypoparathyroidism type 1a and related disorders. *Human Mut.* 2015;36:11. [PMID: 25219572]

Mantovani G, Spada A, Elli FM. Pseudohypoparathyroidism and Gsα-cAMP-linked disorders: current view and open issues. *Nat Rev Endocrinol.* 2016;12:347. [PMID: 27109785]

Shoback DM, Bilezikian JP, Costa AG, et al. Presentation of hypoparathyroidism: etiologies and clinical features. *J Clin Endocrinol Metab.* 2016;101:2300. [PMID: 26943721]

Bone Anatomy and Remodeling

Florencio-Silva F, Sasso GR, Sasso-Cerri E, Simoes MJ, Cerri PS. Biology of bone tissue: structure, function, and factors that influence bone cells. *Bio Med Res International.* 2015;2015:421746. [PMID: 26247020]

Iniguez-Ariza NM, Clarke BL. Bone biology, signaling pathways, and therapeutic targets for osteoporosis. *Maturitas.* 2015;82:245. [PMID: 26255682]

Osteoporosis

Adler RA, El-Hajj Fuleihan G, Bauer DC, et al. Managing osteoporosis in patients on long-term bisphosphonate treatment: report of a task force of the American Society for Bone and Mineral Research. *J Bone Miner Res.* 2016;31:16. [PMID: 26350171]

Bauer DC. Clinical practice. Calcium supplements and fracture prevention. *N Engl J Med.* 2013;369:1537. [PMID: 24131178]

Bischoff-Ferrari HA, Willett WC, Orav EJ, et al. A pooled analysis of vitamin D dose requirements for fracture prevention. *N Engl J Med.* 2012;367:40. [PMID: 22762317]

Black DM, Rosen CJ. Clinical practice. Postmenopausal osteoporosis. *N Engl J Med.* 2016;374:254. [PMID: 26789873]

Cauley JA, Robbins J, Chen Z, et al. Effects of estrogen plus progestin on risk of fracture and bone mineral density: the Women's Health Initiative randomized trial. *JAMA.* 2003;290:1729. [PMID 14519707]

Cosman F. Anabolic and antiresorptive therapy for osteoporosis: combination and sequential approaches. *Curr Osteoporos Rep.* 2014;12:385. [PMID: 25341476]

Cosman F, de Beur SF, LeBoff MS, et al. National Osteoporosis Foundation. Clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int.* 2014;25:2359. [PMID: 25182228]

Cummings SR, San Martin J, McClung MR, et al. Denosumab for prevention of fracture in postmenopausal women with osteoporosis. *N Engl J Med.* 2009;361:756. [PMID: 19671655]

Grossman JM, Gordon R, Ranganath VK, et al. American College of Rheumatology 2010 recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis. *Arthritis Care Res (Hoboken).* 2010;62:1515. [PMID: 20662044]

Khosla S, Burr D, Cauley J, et al. Bisphosphonate-associated osteonecrosis of the jaw: report of a Task Force of the American Society for Bone and Mineral Research. *J Bone Min Res.* 2007;22:1479. [PMID: 176663640]

Lyles KW, Colón-Emeric CS, Magaziner JS, et al. Zoledronic acid in reducing clinical fracture and mortality after hip fracture. *N Engl J Med.* 2007;357:1799. [PMID: 18427590]

Neer RM, Arnaud CD, Zanchetta JR, et al. Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med.* 2001;344:1434. [PMID: 11346808]

Shane E, Burr D, Abrahamsen B, et al. Atypical subtrochanteric and diaphyseal femoral fractures: second report of a task force of the American Society for Bone and Mineral Research. *J Bone Miner Res.* 2014;29:1. [PMID: 23712442]

US Preventive Services Task Force. Screening for osteoporosis: US Preventive Services Task Force recommendation statement. *Ann Intern Med.* 2011;154:356. [PMID: 21242341]

Watts NB, Adler RA, Bilezikian JP, et al. Osteoporosis in men: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2012;97:1802. [PMID: 22675062]

Osteomalacia, Rickets, and Hypophosphatemia

Christov M, Jüppner H. Insights from genetic disorders of phosphate homeostasis. *Semin Nephrol.* 2013;33:143. [PMID: 23465501]

Eguchi H, Tsuruta M, Tani J, Kuwahara R, Hiromatsu Y. Hypophosphatemic osteomalacia due to drug-induced Fanconi's syndrome associated with adefovir dipivoxil treatment for hepatitis B. *Intern Med.* 2014;53:233. [PMID: 24492692]

Feng JQ, Clinkenbeard EL, Yuan B, White KE, Drezner MK. Osteocyte regulation of phosphate homeostasis and bone mineralization underlies the pathophysiology of the heritable disorders of rickets and osteomalacia. *Bone.* 2013;54:213. [PMID: 23403405]

Hautmann AH, Hautmann MG, Kölbl O, Herr W, Fleck M. Tumor-induced osteomalacia: an up-to-date review. *Curr Rheumatol Rep.* 2015;17:512. [PMID: 25900190]

Munns CF, Shaw N, Kiely M, et al. Global consensus recommendations on prevention and management of nutritional rickets. *J Clin Endocrinol Metab.* 2016;101:394. [PMID: 26745253]

FGF23 and Alpha Klotho

Brown RB, Razzaque MS. Dysregulation of phosphate metabolism and conditions associated with phosphate toxicity. *Bonekey Rep.* 2015;4:705. [PMID: 26131357]

Fathi I, Sakr M. Review of tumoral calcinosis: a rare clinico-pathological entity. *World J Clin Cases.* 2014;16:409. [PMID: 25232542]

Fukumoto S. Anti-fibroblast growth factor 23 antibody therapy. *Curr Opin Nephrol Hypertens.* 2014;23:346. [PMID: 24848934]

Paget Disease

Bolland MJ, Cundy T. Paget's disease of bone: clinical review and update. *J Clin Pathol.* 2013;66:924. [PMID: 24043712]

Galson DL, Roodman GD. Pathobiology of Paget's disease of bone. *J Bone Metab.* 2014;21:85. [PMID: 25025000]

Reid IR. Pharmacotherapy of Paget's disease of bone. *Expert Opin Pharmacother.* 2012;13:637. [PMID: 22339140]

Valiet M, Ralston SH. Biology and treatment of Paget's disease of bone. *J Cell Biochem.* 2016;117:289. [PMID: 26212817]

Renal Osteodystrophy

Bover J, Ureña P, Brandenburg V, et al. Adynamic bone disease: from bone to vessels in chronic kidney disease. *Semin Nephrol.* 2014;34:626. [PMID: 25498381]

Drüeke TB, Massy ZA. Changing bone patterns with progression of chronic kidney disease. *Kidney Int.* 2016;89:289. [PMID: 26806832]

Goldsmith DJ, Massy ZA, Brandenburg V. The uses and abuses of Vitamin D compounds in chronic kidney disease-mineral bone disease (CKD-MBD). *Semin Nephrol.* 2014;34:660. [PMID: 25498384]

Isakova T, Ix JH, Sprague SM, et al. Rationale and approaches to phosphate and fibroblast growth factor 23 reduction in CKD. *J Am Soc Nephrol.* 2015;26:2328. [PMID: 25967123]

Koizumi M, Komaba H, Fukagawa M. Parathyroid function in chronic kidney disease: role of FGF23-klotho axis. *Contrib Nephrol.* 2013;180:110. [PMID: 23652554]

Seifert ME, Hruska KA. The kidney-vascular-bone axis in chronic kidney disease-mineral bone disorder. *Transplantation.* 2016;100:497. [PMID: 26356179]

This page intentionally left blank

Glucocorticoids and Adrenal Androgens

Ty B. Carroll, MD, David C. Aron, MD, MS,
James W. Findling, MD, and J. Blake Tyrrell, MD

ACTH	Adrenocorticotropic hormone	HSD	Hydroxysteroid dehydrogenase
ADH	Antidiuretic hormone (vasopressin)	HSP	Heat shock protein
AIRE	Autoimmune regulator gene	ICMA	Immunochemiluminometric assay
APS	Autoimmune polyendocrine syndrome	IL	Interleukin
ARMC5	Armadillo-repeat containing 5	IPSS	Inferior petrosal sinus sampling
AVP	Arginine vasopressin	IRMA	Immunoradiometric assay
BMAH	Bilateral macronodular adrenal hyperplasia	LC/MS/MS	Liquid chromatography tandem mass spectroscopy
CBG	Corticosteroid-binding globulin	LH	Luteinizing hormone
CRH	Corticotropin-releasing hormone	PMN	Polymorphonuclear neutrophil
DAX1	Dosage-sensitive sex reversal-adrenal hypoplasia gene product	PRA	Plasma renin activity
DHEA	Dehydroepiandrosterone	PRKACA	Protein kinase A catalytic subunit
DOC	Deoxycorticosterone	PRL	Prolactin
ELISA	Enzyme-linked immunosorbent assay	PTH	Parathyroid hormone
GH	Growth hormone	RANK	Receptor activator of nuclear factor kappa B
GIP	Glucose-dependent insulinotropic peptide/ Gastrointestinal inhibitory polypeptide	SF-1	Steroidogenic factor-1a
GnRH	Gonadotropin-releasing hormone	SHBG	Sex hormone–binding globulin
HLA	Human leukocyte antigen	StAR	Steroidogenic acute regulatory protein
HPA	Hypothalamic-pituitary-adrenal	TRH	Thyrotropin-releasing hormone
HPLC	High-performance liquid chromatography	TSH	Thyroid-stimulating hormone (thyrotropin)
		VLCFA	Very long chain fatty acid

The adrenal cortex produces many steroid hormones of which the most important are cortisol, aldosterone, and the adrenal androgens. Disorders of the adrenal glands lead to classic endocrinopathies such as Cushing syndrome, Addison disease, hyperaldosteronism, and the syndromes of congenital adrenal hyperplasia. This chapter describes the physiology and disorders of the glucocorticoids and the adrenal androgens. Disorders of aldosterone secretion are discussed in Chapter 10 and congenital defects in adrenal hormone biosynthesis in Chapters 10 and 14.

Hirsutism and virilization (which reflect excess androgen action) are discussed in Chapter 13.

Advances in diagnostic procedures have simplified the evaluation of adrenocortical disorders; in particular, the assay of plasma glucocorticoids, androgens, and adrenocorticotropin (ACTH) has allowed more rapid and precise diagnosis. Recent discoveries have also begun to elucidate the genetic basis for many disorders of the adrenal cortex, and advances in surgical and medical treatment have improved the outlook for patients with these disorders.

EMBRYOLOGY AND ANATOMY

Embryology

The adrenal cortex is of mesodermal origin and derives from a single cell lineage characterized by expression of certain transcription factors such as steroidogenic factor-1 (SF-1). At 2 months' gestation, the cortex, already identifiable as a separate organ, is composed of a fetal zone and a definitive zone similar to the adult adrenal cortex. The adrenal cortex then increases rapidly in size; at mid gestation, it is considerably larger than the kidney and much larger than the adult gland in relation to total body mass. The fetal zone makes up the bulk of the weight of the adrenal cortex at this time. Several genes encoding transcription factors are important in adrenal development and differentiation. These include SF-1 and the product of the dosage-sensitive sex reversal–adrenal hypoplasia gene (*DAX1*), among others; mutations of the *DAX1* gene are associated with congenital adrenal hypoplasia.

The fetal adrenal is under the control of ACTH by mid pregnancy, but the fetal zone is deficient in the activity of 3β-hydroxysteroid dehydrogenase (see section on Biosynthesis of Cortisol and Adrenal Androgens, later) and thus produces mainly dehydroepiandrosterone (DHEA) and DHEA sulfate, which serve as precursors of maternal-placental estrogen production after conversion in the liver to 16α-hydroxylated derivatives. The definitive zone synthesizes a number of steroids and is the major site of fetal cortisol synthesis.

Anatomy

The anatomic relationship of the fetal and definitive zones is maintained until birth, at which time the fetal zone gradually disappears, with a consequent decrease in adrenocortical weight in the 3 months following delivery. During the next 3 years, the adult adrenal cortex develops from cells of the outer layer of the cortex and differentiates into the three adult zones: glomerulosa, fasciculata, and reticularis–adrenal zonation.

The adult adrenal glands, with a combined weight of 8 to 10 g, lie in the retroperitoneum above or medial to the upper poles of the kidneys (Figure 9–1). A fibrous capsule surrounds the gland; the cortex comprises 90% of the adrenal weight, the inner medulla is about 10%.

The adrenal cortex is richly vascularized and receives its main arterial supply from branches of the inferior phrenic artery, the renal arteries, and the aorta. These small arteries form an arterial plexus beneath the capsule and then enter a sinusoidal system that penetrates the cortex and medulla, draining into a single central vein in each gland. The right adrenal vein drains directly into the posterior aspect of the vena cava; the left adrenal vein enters the left renal vein. These anatomic features account for the clinically important fact that it is relatively easier to catheterize the left adrenal vein than it is to catheterize the right adrenal vein.

Microscopic Anatomy

Histologically, the adult cortex is composed of three zones: an outer zona glomerulosa, a zona fasciculata, and an inner zona reticularis (Figure 9–2). However, the inner two zones appear to function as a unit (see later). The zona glomerulosa, which produces aldosterone and constitutes about 15% of adult cortical volume, is deficient in 17α-hydroxylase activity and thus cannot produce cortisol or androgens (see later and Chapter 10). The zona glomerulosa lacks a well-defined structure, and the small lipid-poor cells are scattered beneath the adrenal capsule.
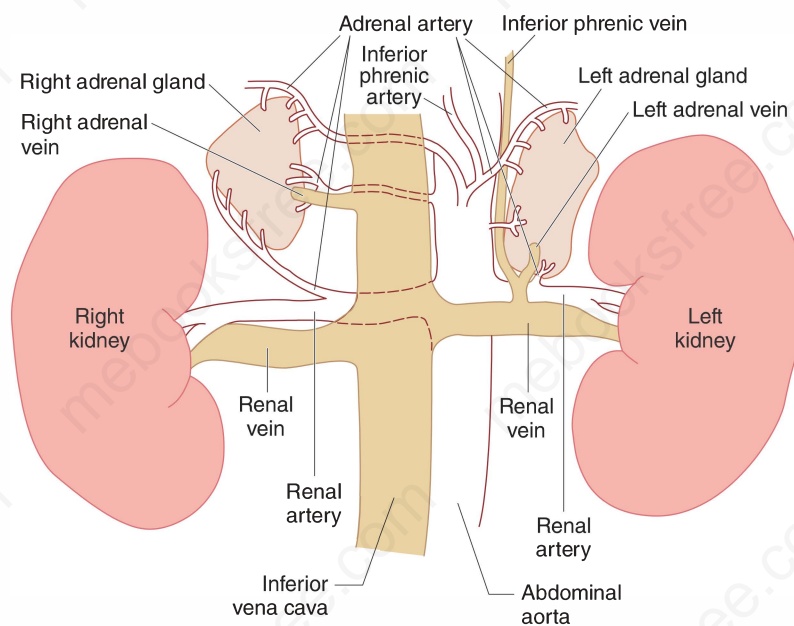
FIGURE 9–1 Location and blood supply of the adrenal glands (schematic). (Reproduced with permission from Felig P, et al, eds. *Endocrinology and Metabolism*, 4th ed. New York: McGraw-Hill Education; 2002.)
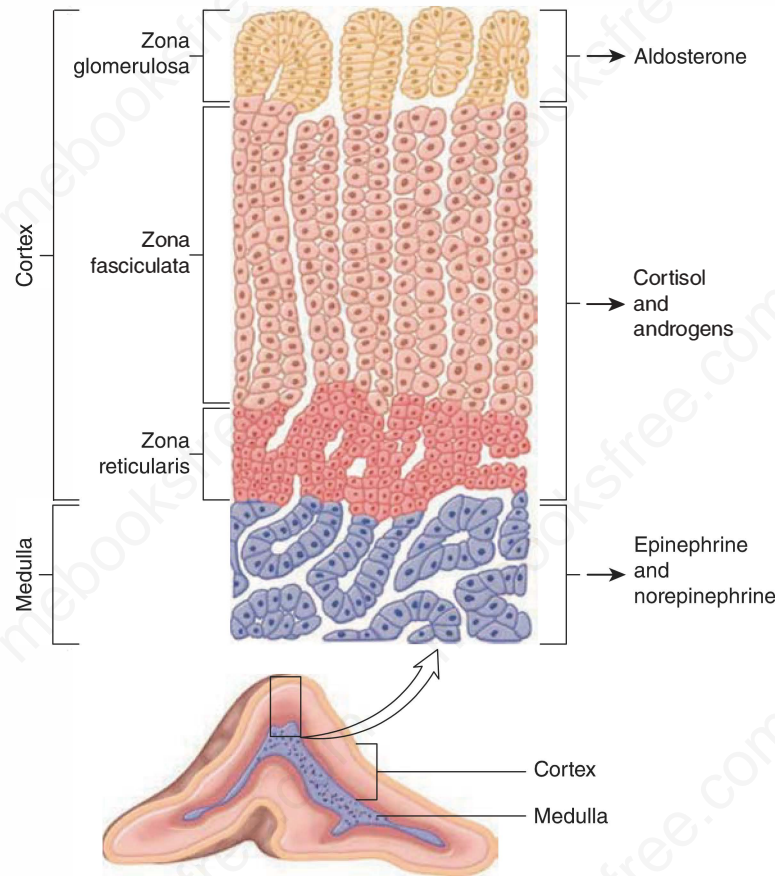
FIGURE 9–2 Zones of the adrenal cortex and medulla. Each of the zones of the cortex (reticularis, fasciculata, and glomerulosa) is derived from different embryologic origins, and their cells have varying hormonal productive capabilities. (Reproduced with permission from Widmaier EP, Raff HR, Strang KT. *Vander's Human Physiology: The Mechanisms of Body Function*. 11th ed. New York: McGraw-Hill Education; 2008.)

The zona fasciculata is the thickest layer of the adrenal cortex, making up about 75% of the cortex, and produces cortisol and androgens. The cells of the zona fasciculata are larger and contain more lipids and thus are termed **clear cells**. These cells extend in columns from the narrow zona reticularis to either the zona glomerulosa or to the capsule. The inner zona reticularis surrounds the medulla and also produces cortisol and androgens. The compact cells of this narrow zone lack significant lipid content but do contain lipofuscin granules. The zonae fasciculata and reticularis are regulated by ACTH; excess or deficiency of this hormone alters their structure and function. Thus, both zones atrophy when ACTH is deficient; when ACTH is present in excess, hyperplasia and hypertrophy of these zones occur. In addition, chronic stimulation with ACTH leads to a gradual depletion of the lipid from the clear cells of the zona fasciculata at the junction of the two zones; these cells thus attain the characteristic appearance of the compact reticularis cells. With chronic excessive stimulation, the compact reticularis cells extend outward and may reach the outer capsule. It is postulated that the zona fasciculata cells can respond acutely to ACTH stimulation with increased cortisol production, whereas the reticularis cells maintain basal glucocorticoid secretion and that induced by prolonged ACTH stimulation.

BIOSYNTHESIS OF CORTISOL AND ADRENAL ANDROGENS

Steroidogenesis

The major hormones secreted by the adrenal cortex are cortisol, the androgens, and aldosterone. The carbon atoms in the steroid molecule are numbered as shown in Figure 9–3, and the major biosynthetic pathways and hormonal intermediates are illustrated in Figure 9–4.

The scheme of adrenal steroidogenic synthesis has been clarified by analysis of the steroidogenic enzymes. Most of these enzymes belong to the family of cytochrome P450 oxygenases (see Table 9–1 for current and historical nomenclature conventions). In mitochondria, the *CYP11A* gene, located on chromosome 15, encodes P450scc, the enzyme responsible for cholesterol side chain cleavage. *CYP11B1*, a gene located on chromosome 8, encodes P450c11, another mitochondrial enzyme, which mediates 11β-hydroxylation in the zona reticularis and zona fasciculata. This reaction converts 11-deoxycortisol to cortisol and 11-deoxycorticosterone (11-DOC) to corticosterone. In the zona glomerulosa, *CYP11B2*, also located on chromosome 8, encodes the enzyme P450aldo, also known as aldosterone synthase.
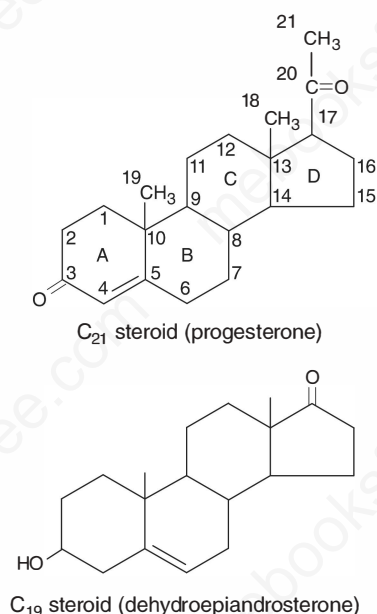
FIGURE 9–3 Structure of adrenocortical steroids. The letters in the formula for progesterone identify the A, B, C, and D rings; the numbers show the positions in the basic C_{21} steroid structure. The angular methyl groups (positions 18 and 19) are usually indicated simply by straight lines, as in the lower formula. Dehydroepiandrosterone is a 17-ketosteroid formed by cleavage of the side chain of the C_{21} steroid 17-hydroxypregnenolone and its replacement by an O atom. (Reproduced with permission from Ganong WF. *Review of Medical Physiology*. 14th ed. McGraw-Hill; 1989.)

P450aldo mediates 11β-hydroxylation, 18-hydroxylation, and 18-oxidation to convert 11-DOC to corticosterone, 18-hydroxycorticosterone, and aldosterone, respectively. In the endoplasmic reticulum, gene *CYP17*, located on chromosome 10, encodes a single enzyme, P450c17, which mediates both 17α-hydroxylase activity and 17,20-lyase activity, and the gene *CYP21A2* encodes the P450c21, which mediates 21-hydroxylation of both progesterone and 17-hydroxyprogesterone. The 3β-hydroxysteroid dehydrogenase: $\Delta^{5,4}$-isomerase activities are mediated by a single non-P450 microsomal enzyme (see Figure 9–4).

A. Zones and steroidogenesis Because of enzymatic differences between the zona glomerulosa and the inner two zones, the adrenal cortex functions as two separate units, with differing regulation and secretory products. Thus, the zona glomerulosa, which produces aldosterone, lacks 17α-hydroxylase activity and cannot synthesize 17α-hydroxypregnenolone and 17α-hydroxyprogesterone, which are the precursors of cortisol and the adrenal androgens. The synthesis of aldosterone by this zone is primarily regulated by the renin-angiotensin system and by potassium (see Chapter 10).

The zona fasciculata and zona reticularis (see Figure 9–4) produce cortisol, androgens, and small amounts of estrogens. These zones, primarily regulated by ACTH, do not express the gene *CYP11B2* (encoding P450aldo) and therefore cannot convert 11-DOC to aldosterone (see Chapter 10).
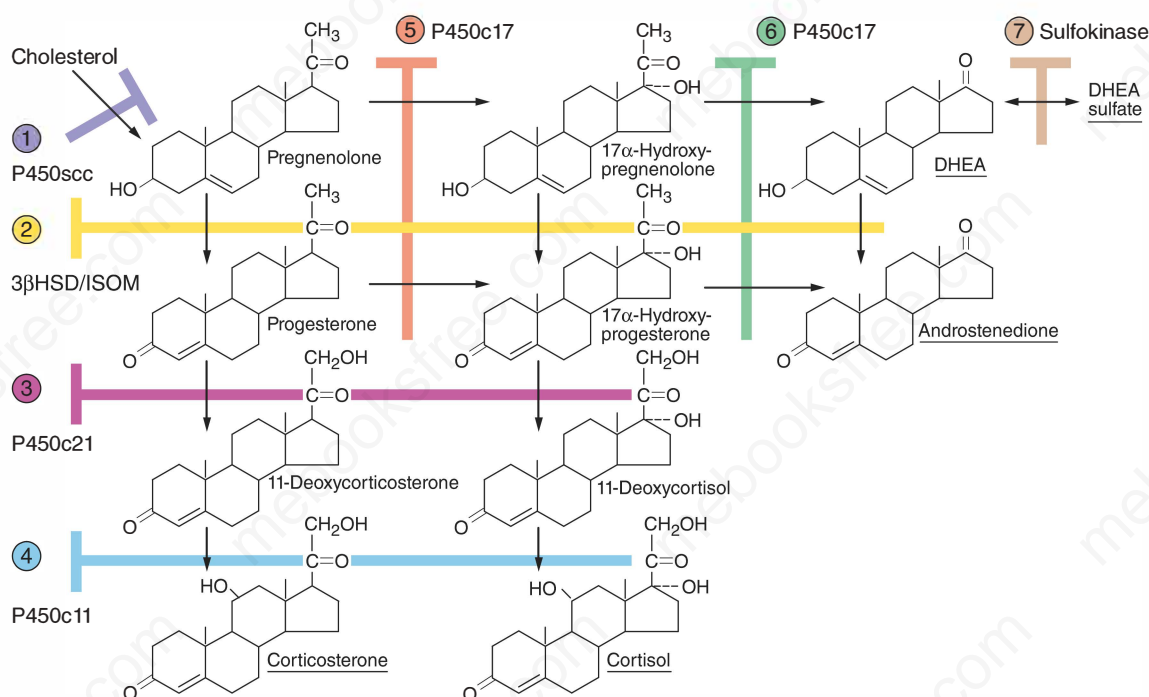
FIGURE 9–4 Steroid biosynthesis in the zona fasciculata and zona reticularis of the adrenal cortex. The major secretory products are underlined. The enzymes for the reactions are numbered on the left and at the top of the chart, with the steps catalyzed shown by the shaded bars. ① P450scc, cholesterol 20,22-hydroxylase:20,22 desmolase activity; ② 3βHSD/ISOM, 3-hydroxysteroid dehydrogenase; δ⁵-oxosteroid isomerase activity; ③ P450c21, 21α-hydroxylase activity; ④ P450c11, 11β-hydroxylase activity; ⑤ P450c17, 17α-hydroxylase activity; ⑥ P450c17, 17,20-lyase/desmolase activity; ⑦ sulfokinase. (Modified with permission from Ganong WF. *Review of Medical Physiology*. 16th ed. McGraw-Hill; 1993.)

TABLE 9–1   The main components of the steroidogenic pathway.

Enzyme	Gene	Chromosomal Location	Enzyme Activity (or Activities)	Subcellular Localization	Characteristic Features of Normal Tissue-Specific Expression
Steroidogenic acute regulatory protein (StAR)	StAR	8p11.2	Activation of peripheral type benzodiazepine receptor	Outer mitochondrial membrane	All steroid hormone–producing cells except the placenta, Schwann cells, and the brain
Peripheral type benzodiazepine receptor	PDA	22q13.31	Regulated cholesterol channel	Forms channel at the contact sites between the outer and inner mitochondrial membranes	All steroid hormone–producing cells
P450scc	CYP11A	15q23-24	Cholesterol-20,22-desmolase	Matrix side of inner mitochondrial membrane	All steroid hormone–producing cells
3β-Hydroxysteroid dehydrogenase (HSD)/isomerase	3β-Hydroxysteroid dehydrogenase type 1	1p13, the HSD3B1 and HSD3B2 loci are located 1-2 cm from the centromeric marker D1Z5	3β-Hydroxysteroid dehydrogenase	Smooth endoplasmic reticulum	Expressed in syncytiotrophoblast cells, sebaceous glands
	3β-Hydroxysteroid dehydrogenase type 2				Expressed in the definitive adrenal cortex and the gonads. Absent from the fetal zone of the adrenal cortex
P450c21	CYP21B	6p21.3 (close to HLA locus)	21-hydroxylase	Smooth endoplasmic reticulum	Only in adrenal cortex (all zones); low levels in fetal zone
	CYP21A	6p21.3 (close to HLA locus)	Pseudogene	N/A	N/A
P450c11	CYP11B1	8q21-22	11β-hydroxylase	Matrix side of inner mitochondrial membrane	Only expressed in zona fasciculata and zona reticularis of the adrenal cortex; low levels in fetal zone
P450aldo	CYP11B2	8q24.3	Aldosterone synthase: 11β-hydroxylase, 18-hydroxylase, 18-oxidase	Matrix side of inner mitochondrial membrane	Only expressed in the zona glomerulosa of the adrenal cortex; low levels in fetal zone
P450c17	CYP17	10q24-25	17α-hydroxylase, 17,20-lyase	Smooth endoplasmic reticulum	Absent from the zona glomerulosa in adults, the placenta, and the definitive zone of fetal adrenal cortex until the third trimester
P450arom	CYP19	15q	Aromatase	Smooth endoplasmic reticulum	Expressed in numerous tissues including gonads, brain, adrenal, adipose tissue, and bone

Modified with permission from Kacsoh B. *Endocrine Physiology*. New York: McGraw-Hill Education; 2000.

B. Cholesterol uptake and synthesis   Synthesis of cortisol and the androgens by the zonae fasciculata and reticularis begins with cholesterol, as does the synthesis of all steroid hormones. Plasma lipoproteins are the major source of adrenal cholesterol, although synthesis within the gland from acetate also occurs. Low-density lipoprotein accounts for about 80% of cholesterol delivered to the adrenal gland. A small pool of free cholesterol within the adrenal is available for rapid synthesis of steroids when the adrenal is stimulated. When stimulation occurs, there is also increased hydrolysis of stored cholesteryl esters to free cholesterol, increased uptake from plasma lipoproteins, and increased cholesterol synthesis within the gland. The acute

response to a steroidogenic stimulus is mediated by the steroidogenic acute regulatory protein (StAR). This mitochondrial phosphoprotein enhances cholesterol transport from the outer to the inner mitochondrial membrane. Mutations in the *StAR* gene result in congenital lipoid adrenal hyperplasia with severe cortisol and aldosterone deficiencies at birth.

C. Cholesterol metabolism   The conversion of cholesterol to pregnenolone is the rate-limiting step in adrenal steroidogenesis and the major site of ACTH action in the adrenal. This step occurs in the mitochondria and involves two hydroxylations and then the side chain cleavage of cholesterol. A single enzyme, CYP11A, mediates this process; each step requires molecular oxygen and a pair of electrons. The latter are donated by nicotinamide adenine dinucleotide phosphate (NADPH) to adrenodoxin reductase (ferredoxin reductase), a flavoprotein, and then to adrenodoxin, an iron-sulfur protein, and finally to CYP11A. Both adrenodoxin reductase and adrenodoxin are also involved in the action of CYP11B1 (see earlier). Electron transport to microsomal cytochrome P450 involves P450 reductase, a flavoprotein distinct from adrenodoxin reductase. Pregnenolone is then transported outside the mitochondria before further steroid synthesis occurs.

D. Synthesis of cortisol   Cortisol synthesis proceeds by 17α-hydroxylation of pregnenolone by CYP17 within the smooth endoplasmic reticulum to form 17α-hydroxypregnenolone. This steroid is then converted to 17α-hydroxyprogesterone after conversion of its 5,6 double bond to a 4,5 double bond by the 3β-hydroxysteroid dehydrogenase: $\Delta^{5,4}$-oxosteroid isomerase enzyme complex, which is also located within the smooth endoplasmic reticulum. An alternative but apparently less important pathway in the zonae fasciculata and reticularis is from pregnenolone to progesterone to 17α-hydroxyprogesterone (see Figure 9–4).

The next step, which is again microsomal, involves the 21-hydroxylation by CYP21A2 of 17α-hydroxyprogesterone to form 11-deoxycortisol; this compound is further hydroxylated within mitochondria by 11β-hydroxylation (CYP11B1) to form cortisol. The zona fasciculata and zona reticularis also produce 11-DOC, 18-hydroxydeoxycorticosterone, and corticosterone. However, as noted earlier, the absence of the mitochondrial enzyme CYP11B2 prevents production of aldosterone by these zones of the adrenal cortex. Cortisol secretion under basal (ie, nonstressed) conditions ranges from 8 to 25 mg/d (22-69 μmol/d), with a mean of about 9.2 mg/d (25 μmol/d)—rates lower than most previous calculations.

E. Synthesis of androgens   The production of adrenal androgens from pregnenolone and progesterone requires prior 17α-hydroxylation (CYP17) and thus does not occur in the zona glomerulosa. The major quantitative production of androgens is by conversion of 17α-hydroxypregnenolone to the 19-carbon compounds (C-19 steroids) DHEA and its sulfate conjugate DHEA sulfate. Thus, 17α-hydroxypregnenolone

undergoes removal of its two-carbon side chain at the C_{17} position by microsomal 17,20-desmolase (CYP17), yielding DHEA with a keto group at C_{17}. DHEA is then converted to DHEA sulfate by a reversible adrenal sulfokinase. The other major adrenal androgen, androstenedione, is produced mostly from DHEA, mediated by CYP17, and possibly from 17α-hydroxyprogesterone, also by CYP17. Androstenedione can be converted to testosterone, although adrenal secretion of this hormone is minimal. The adrenal androgens, DHEA, DHEA sulfate, and androstenedione, have minimal intrinsic androgenic activity, and they contribute to androgenicity by their peripheral conversion to the more potent androgens testosterone and dihydrotestosterone. Although DHEA and DHEA sulfate are secreted in greater quantities, androstenedione is qualitatively more important, because it is more readily converted peripherally to testosterone (see Chapter 12). Of note, studies have identified *de novo* synthesis of some steroid hormones in nerve and cardiac tissues, where they appear to act as paracrine or autocrine factors. Steroidogenic enzymes (eg, 3β-hydroxysteroid dehydrogenase and aromatase) are expressed in many tissues.

Regulation of Secretion

A. Secretion of CRH and ACTH   ACTH is the trophic hormone of the zonae fasciculata and reticularis and the major regulator of cortisol and adrenal androgen production, although other factors produced within the adrenal, including neurotransmitters, neuropeptides, and nitric oxide also play a role. ACTH in turn is regulated by the hypothalamus and central nervous system via neurotransmitters and corticotropin-releasing hormone (CRH) and arginine vasopressin (AVP). The neuroendocrine control of CRH and ACTH secretion involves three mechanisms (see later and Chapter 4).

B. ACTH effects on the adrenal cortex   ACTH administration leads to the rapid synthesis and secretion of steroids; plasma levels of these hormones rise within minutes. ACTH increases RNA, DNA, and protein synthesis. Chronic stimulation leads to adrenocortical hyperplasia and hypertrophy; conversely, ACTH deficiency results in decreased steroidogenesis and is accompanied by adrenocortical atrophy, decreased gland weight, and decreased protein and nucleic acid content.

C. ACTH and steroidogenesis   ACTH binds to high-affinity plasma membrane receptors, thereby activating adenylyl cyclase and increasing cyclic adenosine monophosphate, which in turn activates intracellular phosphoprotein kinases (Figure 9–5), including StAR. Mutations of the ACTH receptor (also known as the *melanocortin-2 receptor*) gene are associated with familial glucocorticoid deficiency. ACTH action results in increased free cholesterol formation as a consequence of increased cholesterol esterase activity and decreased cholesteryl ester synthetase as well as increased lipoprotein uptake by the adrenal cortex. This process stimulates the rate-limiting
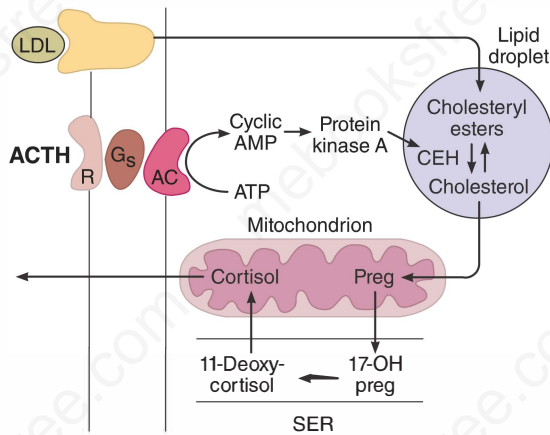
FIGURE 9–5 Mechanism of action of ACTH on cortisol-secreting cells in the inner two zones of the adrenal cortex. When ACTH binds to its receptor (R), adenylyl cyclase (AC) is activated via G_s. The resulting increase in cAMP activates protein kinase A, and the kinase phosphorylates cholesteryl ester hydrolase (CEH), increasing its activity. Consequently, more free cholesterol is formed and converted to pregnenolone in the mitochondria. Note that in the subsequent steps in steroid biosynthesis, products are shuttled between the mitochondria and the smooth endoplasmic reticulum (SER). (Reproduced with permission from Ganong WF. *Review of Medical Physiology*. 16th ed. McGraw-Hill; 1993.)

step—cholesterol delivery to the side chain cleavage enzyme (P450scc or CYP11A1) for conversion to Δ^5-pregnenolone, thereby initiating steroidogenesis.

D. Neuroendocrine control Cortisol secretion is closely regulated by ACTH, and plasma cortisol levels parallel those of ACTH (Figure 9–6). There are three mechanisms of neuroendocrine control: (1) episodic secretion and the circadian rhythm of
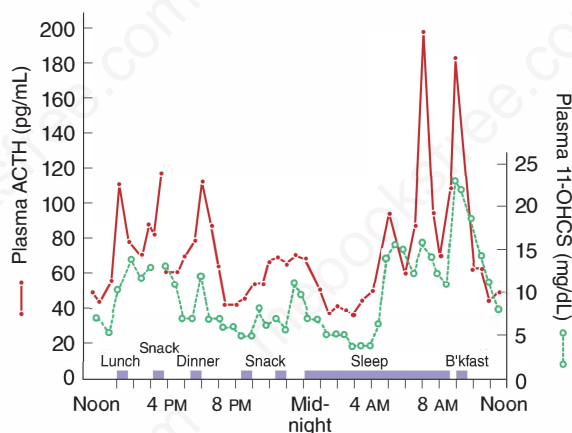
FIGURE 9–6 Fluctuations in plasma ACTH and glucocorticoids (11-OHCS) throughout the day. Note the greater ACTH and glucocorticoid rises in the morning before awakening. (Reproduced with permission from Krieger DT et al. Characterization of the normal temporal pattern of corticosteroid levels. *J Clin Endocrinol Metab*. 1971 Feb;32(2):266-284.)

ACTH, (2) stress responsiveness of the hypothalamic-pituitary-adrenal (HPA) axis, and (3) feedback inhibition by cortisol of ACTH secretion.

1. **Circadian rhythm**—Circadian rhythm is superimposed on episodic secretion; it is the result of central nervous system events that regulate both the number and magnitude of CRH and ACTH secretory episodes. Cortisol secretion is low in the late evening and continues to decline in the first several hours of sleep, at which time plasma cortisol levels may be nearly undetectable. During the third and fifth hours of sleep there is an increase in secretion; but the major secretory episodes begin in the sixth to eighth hours of sleep (see Figure 9–6) and then begin to decline as wakefulness occurs. About half of the total daily cortisol output is secreted during this period. Cortisol secretion then gradually declines during the day, with fewer secretory episodes of decreased magnitude; however, there is increased cortisol secretion in response to eating and exercise.

    Although this general pattern is consistent, there is considerable intra- and inter-individual variability, and the circadian rhythm may be altered by changes in sleep pattern, light-dark exposure, and feeding times. The rhythm is also changed by (1) physical stresses such as major illness, surgery, trauma, or starvation; (2) psychologic stress, including severe anxiety, endogenous depression, and the manic phase of manic-depressive psychosis; (3) central nervous system and pituitary disorders; (4) Cushing syndrome; (5) liver disease and other conditions that affect cortisol metabolism; (6) chronic renal failure; and (7) alcoholism. Cyproheptadine inhibits the circadian rhythm, possibly by its antiserotonergic effects, whereas other drugs usually have no effect.

2. **Stress responsiveness**—Plasma ACTH and cortisol secretion are also characteristically responsive to physical stress. Thus, plasma ACTH and cortisol are secreted within minutes following the onset of stresses such as surgery and hypoglycemia, and these responses abolish circadian periodicity if the stress is prolonged. Stress responses originate in the central nervous system and increase hypothalamic CRH and thus pituitary ACTH secretion. Stress responsiveness of plasma ACTH and cortisol is abolished by prior high-dose glucocorticoid administration and in spontaneous Cushing syndrome; conversely, the responsiveness of ACTH secretion is enhanced following adrenalectomy. Regulation of the HPA axis is linked to that of the immune system. For example, interleukin-1 (IL-1) stimulates ACTH secretion, and cortisol inhibits IL-1 synthesis.

3. **Feedback inhibition**—The third major regulator of ACTH and cortisol secretion is that of feedback inhibition by glucocorticoids of CRH, ACTH, and cortisol secretion. Glucocorticoid feedback inhibition occurs at both the pituitary and hypothalamus and involves two distinct mechanisms—fast and delayed feedback inhibition.

    Fast feedback inhibition of ACTH secretion is rate-dependent; that is, it depends on the rate of increase of the glucocorticoid but not the dose administered. This phase is rapid (within minutes) and transient (lasting <10 minutes), suggesting mediation by a noncytosolic glucocorticoid receptor mechanism. **Delayed feedback inhibition** is both time- and dose-dependent. With continued glucocorticoid administration, ACTH levels continue to decrease and become unresponsive to stimulation, ultimately resulting in suppression of CRH and ACTH release and

atrophy of the zonae fasciculata and reticularis. The suppressed HPA axis fails to respond to stress and stimulation. Delayed feedback appears to act via the classic glucocorticoid receptor mechanism (see later).

E. ACTH effects on regulation of androgen production Adrenal androgen production in adults is also regulated by ACTH; both DHEA and androstenedione exhibit circadian periodicity in concert with ACTH and cortisol. In addition, plasma concentrations of DHEA and androstenedione increase rapidly with ACTH administration and are suppressed by glucocorticoid administration. DHEA sulfate, because of its slow metabolic clearance rate, does not exhibit a diurnal rhythm. The existence of a separate anterior pituitary hormone that regulates adrenal androgen secretion has long been postulated but not yet proved.

CIRCULATION OF CORTISOL AND ADRENAL ANDROGENS

Cortisol and the adrenal androgens circulate bound to plasma proteins. The plasma half-life of cortisol (60-90 minutes) is determined by the extent of plasma binding and by the rate of metabolic inactivation.

Plasma-Binding Proteins

Cortisol and adrenal androgens are secreted in an unbound state; however, these hormones bind to plasma proteins on entering the circulation. Cortisol binds mainly to corticosteroid-binding globulin (CBG, transcortin) and to a lesser extent to albumin, whereas the androgens bind chiefly to albumin. Bound steroids are biologically inactive; the unbound or free fraction is active. The plasma proteins may provide a pool of circulating cortisol by delaying metabolic clearance, thus preventing more marked fluctuations of plasma-free cortisol levels during episodic secretion by the gland. Because there are no binding proteins in saliva, salivary cortisol reflects free cortisol.

Free and Bound Cortisol

Under basal conditions, about 10% of the circulating cortisol is free, about 75% is bound to CBG, and the remainder is bound to albumin. The plasma free cortisol level is approximately 1 µg/dL, and it is this biologically active cortisol that is regulated by ACTH.

A. Corticosteroid-binding globulin Corticosteroid-binding globulin (CBG) has a molecular weight of about 50,000, is produced by the liver, and binds cortisol with high affinity. The CBG in plasma has a cortisol-binding capacity of about 25 µg/dL. When total plasma cortisol concentrations rise above this level, the free concentration rapidly increases and exceeds its usual fraction of 10% of the total cortisol. Other endogenous steroids usually do not appreciably affect cortisol binding to CBG; an exception is in late pregnancy, when progesterone may occupy

about 25% of the binding sites on CBG. Synthetic steroids do not bind significantly to CBG—with the exception of prednisolone. CBG levels are increased in high-estrogen states (pregnancy; estrogen or oral contraceptive use), hyperthyroidism, diabetes, certain hematologic disorders, and on a genetic basis. CBG concentrations are decreased in familial CBG deficiency, hypothyroidism, and protein deficiency states such as severe liver disease or nephrotic syndrome.

B. Albumin Albumin has a much greater capacity for cortisol binding but a lower affinity. It normally binds about 15% of the circulating cortisol, and this proportion increases when the total cortisol concentration exceeds the CBG-binding capacity. Synthetic glucocorticoids are extensively bound to albumin (eg, about 75% of dexamethasone in plasma is bound to albumin).

C. Androgen binding Androstenedione, DHEA, and DHEA sulfate circulate weakly bound to albumin. However, testosterone is bound extensively to a specific globulin, sex hormone–binding globulin (SHBG) (see Chapter 12).

METABOLISM OF CORTISOL AND ADRENAL ANDROGENS

The metabolism of these steroids renders them inactive and increases their water solubility, as does their subsequent conjugation with glucuronide or sulfate groups. These inactive, conjugated metabolites are more readily excreted by the kidney. The liver is the major site of steroid catabolism and conjugation, and 90% of these metabolized steroids are excreted by the kidney.

Conversion and Excretion of Cortisol

Cortisol is modified extensively before excretion in urine; less than 1% of secreted cortisol appears in the urine unchanged.

A. Hepatic conversion Hepatic metabolism of cortisol involves a number of metabolic conversions of which the most important (quantitatively) is the irreversible inactivation of the steroid by Δ^4-reductases, which reduce the 4,5 double bond of the A ring. Dihydrocortisol, the product of this reaction, is then converted to tetrahydrocortisol by a 3-hydroxysteroid dehydrogenase. Cortisol is also converted extensively by 11β-hydroxysteroid dehydrogenase to the biologically inactive cortisone, which is then metabolized by the enzymes described earlier to yield tetrahydrocortisone. Tetrahydrocortisol and tetrahydrocortisone can be further altered to form the cortoic acids. These conversions result in the excretion of approximately equal amounts of cortisol and cortisone metabolites. Cortisol and cortisone are also metabolized to the cortols and cortolones and to a lesser extent by other pathways (eg, to 6β-hydroxycortisol).

B. Hepatic conjugation More than 95% of cortisol and cortisone metabolites are conjugated by the liver and then

reenter the circulation to be excreted in the urine. Conjugation is mainly with glucuronic acid at the 3α-hydroxyl position.

C. Variations in clearance and metabolism

The metabolism of cortisol is altered by a number of circumstances. It is decreased in infants and in the elderly. It is impaired in chronic liver disease, leading to decreased renal excretion of cortisol metabolites; however, the plasma cortisol level remains normal. Hypothyroidism decreases both metabolism and excretion; conversely, hyperthyroidism accelerates these processes. Cortisol clearance may be reduced in starvation and anorexia nervosa and is also decreased in pregnancy because of the elevated CBG levels. The metabolism of cortisol to 6β-hydroxycortisol is increased in the neonate, in pregnancy, with estrogen therapy, and in patients with liver disease or severe chronic illness. Cortisol metabolism by this pathway is also increased by drugs that induce hepatic microsomal enzymes, including barbiturates, phenytoin, mitotane, aminoglutethimide, and rifampin. These alterations generally are of minor physiologic importance as the free cortisol levels remain relatively stable in these conditions. However, they result in decreased excretion of the urinary metabolites of cortisol measured as 17-hydroxycorticosteroids. These conditions and drugs have a greater influence on the metabolism of synthetic glucocorticoids and may result in

inadequate plasma levels of the administered glucocorticoid because of rapid clearance and metabolism.

D. Cortisol-cortisone shunt

Aldosterone is the principal mineralocorticoid controlling sodium and potassium exchange in the distal nephron. Mineralocorticoid receptors in the kidney are responsible for this effect, and the sensitivities of both the glucocorticoid receptor and the mineralocorticoid receptor for cortisol *in vitro* are similar. Small changes in aldosterone affect sodium and potassium exchange in the kidney, whereas free and biologically active cortisol does not, yet cortisol circulates in much higher concentrations. This apparent paradox is explained by an intracellular enzyme—11β-hydroxysteroid dehydrogenase type 2 (11β-HSD2)—that metabolizes cortisol to inactive cortisone and protects the mineralocorticoid receptor from cortisol binding (Figure 9–7). However, when circulating cortisol is extremely high (as in severe Cushing syndrome), this prereceptor metabolism of cortisol is overwhelmed and the mineralocorticoid receptor is activated by cortisol, resulting in volume expansion, hypertension, and hypokalemia. The active ingredient of licorice (glycyrrhizic acid) actually inhibits 11β-HSD2 and gives cortisol free access to the unprotected mineralocorticoid receptor in the kidney, causing hypokalemia and hypertension. In addition, some tissues can actually convert the inactive
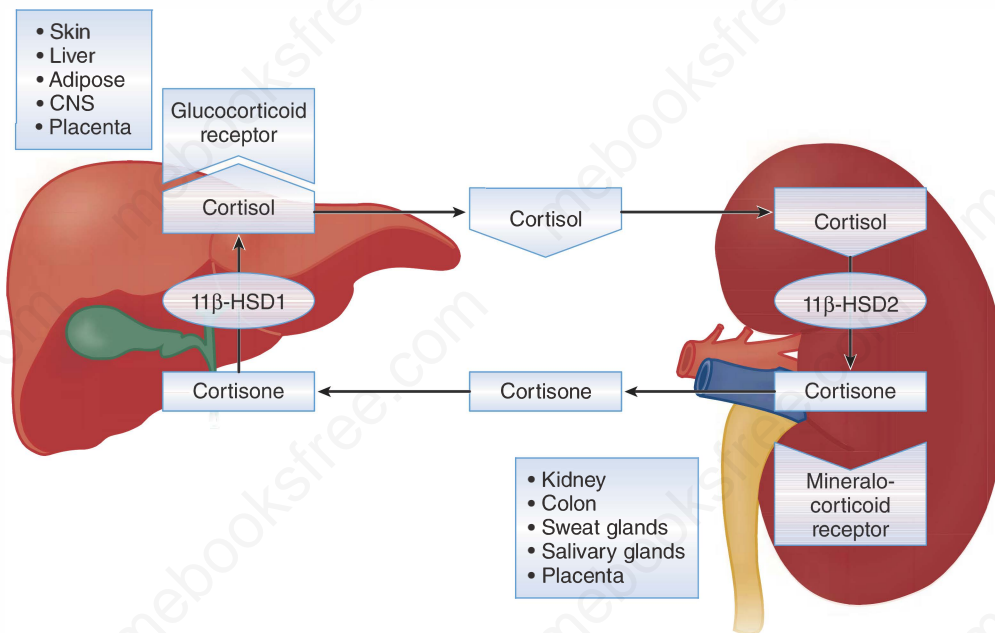
FIGURE 9–7 Cortisol-cortisone shunt. Contrasting functions of the isozymes of 11β-HSD. 11β-HSD2 is an exclusive 11β-dehydrogenase that acts in classical aldosterone target tissues to exclude cortisol from otherwise nonselective mineralocorticoid receptors. Inactivation of cortisol also occurs in placenta. 11β-HSD1 is a predominant 11β-reductase *in vivo* that acts in many tissues to increase local intracellular glucocorticoid concentrations and thereby maintain adequate exposure of relatively low affinity glucocorticoid receptors to their ligand.

cortisone to cortisol with the isoform called 11β-hydroxysteroid dehydrogenase type 1 (11β-HSD1). The skin expresses this enzyme, explaining why cortisone cream can be effective. More importantly, the liver expresses 11β-HSD1 and can activate cortisone to cortisol, thereby completing the cortisol-cortisone shunt such that the kidney inactivates cortisol to cortisone and the liver can reactivate cortisone to cortisol. The expression of 11β-HSD1 in adipose tissue may contribute to abdominal obesity seen in metabolic syndrome without biochemical hypercortisolism.

Conversion and Excretion of Adrenal Androgens

Adrenal androgen metabolism results either in degradation and inactivation or the peripheral conversion of these weak androgens to their more potent derivatives testosterone and dihydrotestosterone. DHEA is readily converted within the adrenal to DHEA sulfate, the adrenal androgen secreted in greatest amount. DHEA secreted by the gland is also converted to DHEA sulfate by the liver and kidney, or it may be converted to Δ^4-androstenedione. DHEA sulfate may be excreted without further metabolism; however, both it and DHEA are also metabolized to 7α- and 16α-hydroxylated derivatives and by 17β reduction to Δ^5-androstenediol and its sulfate. Androstenedione is converted either to testosterone or by reduction of its 4,5 double bond to etiocholanolone or androsterone, which may be further converted by 17α reduction to etiocholanediol and androstanediol, respectively. Testosterone is converted to dihydrotestosterone in androgen-sensitive tissues by 5β reduction, and it, in turn, is mainly metabolized by 3α reduction to androstanediol. The metabolites of these androgens are conjugated either as glucuronides or sulfates and excreted in the urine.

BIOLOGIC EFFECTS OF ADRENAL STEROIDS

Glucocorticoids

Although glucocorticoids were originally so called because of their influence on glucose metabolism, they are currently defined as steroids that exert their effects by binding to specific cytosolic receptors that mediate the actions of these hormones. These glucocorticoid receptors are present in virtually all tissues, and glucocorticoid-receptor interaction is responsible for most of the known effects of these steroids. Alterations in the structure of the glucocorticoids have led to the development of synthetic compounds with greater glucocorticoid activity. The increased activity of these compounds is due to increased affinity for the glucocorticoid receptors and delayed plasma clearance, which increases tissue exposure. In addition, many of these synthetic glucocorticoids have negligible mineralocorticoid effects and thus do not result in sodium retention, hypertension, and hypokalemia. This section describes the molecular mechanisms of glucocorticoid action and the effects on individual metabolic functions and tissues (Table 9–2).

Molecular Mechanisms

Glucocorticoid action is initiated by entry of the steroid into the cell and binding to the cytosolic glucocorticoid–receptor proteins (see Figures 1–13, 1–15, and 1–16). The most abundant cytoplasmic glucocorticoid–receptor complex includes two subunits of the 90-kDa heat shock protein (hsp) 90. After binding, the hsp90 subunits dissociate and activated hormone–receptor complexes enter the nucleus and interact with nuclear chromatin acceptor sites. The DNA-binding domain of the receptor is a cysteine-rich region which, when it chelates zinc, assumes a conformation called a zinc finger. The receptor–glucocorticoid complex acts via two mechanisms: (1) binding to specific sites in nuclear DNA, the glucocorticoid regulatory elements; and (2) interactions with other transcription factors such as nuclear factor κB, an important regulator of cytokine genes. These result in altered expression of specific genes and the transcription of specific mRNAs. The resulting proteins elicit the glucocorticoid response, which may be inhibitory or stimulatory depending on the specific gene and tissue affected. Although glucocorticoid receptors are similar in many tissues, the proteins synthesized in response to glucocorticoids vary widely and are the result of expression of specific genes in different cell types. The mechanisms underlying this specific regulation are unknown. Analyses of cloned complementary DNAs for human glucocorticoid receptors have revealed marked structural and amino acid sequence homology between glucocorticoid receptors and receptors for other steroid hormones (eg, mineralocorticoids, estrogen, progesterone) as well as for thyroid hormone and the oncogene v-*erb*-A. Although the steroid-binding domain of the glucocorticoid receptor confers specificity for glucocorticoid binding, glucocorticoids such as cortisol and corticosterone bind to the mineralocorticoid receptor with an affinity equal to that of aldosterone. Mineralocorticoid receptor specificity is maintained by the expression of 11β-HSD in classic mineralocorticoid-sensitive tissues—the cortisol-cortisone shunt (see earlier).

Although glucocorticoid–receptor complexes and their subsequent regulation of gene expression are responsible for most glucocorticoid effects, other effects may occur through plasma membrane receptors.

Glucocorticoid Agonists and Antagonists

The study of glucocorticoid receptors has led to the definition of glucocorticoid agonists and antagonists. These studies have also identified a number of steroids with mixed effects termed partial agonists, partial antagonists, or partial agonist–partial antagonists. In addition, novel glucocorticoid receptor ligands are being developed that have more selectivity in terms of receptor binding or transcription of specific genes.

A. Agonists   In humans, cortisol, synthetic glucocorticoids (eg, prednisolone, dexamethasone), corticosterone, and aldosterone are glucocorticoid agonists. The synthetic glucocorticoids have substantially higher affinity for the glucocorticoid receptor, and these have greater glucocorticoid activity than cortisol when

TABLE 9–2 The main targets and actions of glucocorticoids and the consequences of Cushing disease and Addison disease.

Target System	Specific Target	Physiologic Function	Cushing Disease	Addison Disease
Intermediary metabolism	Liver	Increased expression of gluconeogenic enzymes, phosphoenolpyruvate car-boxykinase, glucose-6-phosphatase, and fructose-2,6-bisphosphatase	Increased hepatic glucose output; together with insulin, increased hepatic glycogen stores	Diminished hepatic glucose output and glycogen stores
	Adipose tissue	Permissive for lipolytic signals (catechol-amines, GH) leading to elevated plasma FFA to fuel gluconeogenesis	Overall effect (together with insulin): central obesity (truncal obesity, moon facies, and buffalo hump)	Decreased adiposity and decreased lipolysis
	Skeletal muscle	Degradation of fibrillar muscle proteins by activating the ubiquitin pathway, thereby providing amino acid sub-strates for gluconeogenesis	Muscle weakness and wasting mainly in proximal muscles; increased urinary nitrogen excre-tion (urea from amino acids)	Muscle weakness, decreased muscle glycogen stores; decreased urinary nitrogen excretion
	Plasma glucose	Maintains plasma glucose during fasting (anti-hypoglycemic action); increases plasma glucose during stress (hyper-glycemic action)	Impaired glucose tolerance, insulin-resistant diabetes mellitus; increased plasma glucose is due to decreased peripheral glucose utilization and increased hepatic glucose output	Hypoglycemia, increased insulin sensitivity
Calcium homeostasis	Kidney	Decreased reabsorption of calcium	Hypercalciuria without hypercalce-mia leading to secondary hyperparathyroidism	Retardation of bone growth mainly through decreased GH; hypercalcemia possible
	Bone, cartilage	Inhibition of collagen synthesis and bone deposition	Retardation of bone growth and bone age by direct action and by decreas-ing GH; osteoporosis in adults	
	Gastrointestinal tract	Inhibition of calcium, magnesium, and phosphate absorption by antagonizing calcitriol actions		
Other endocrine systems	Hypothalamus, pituitary	Decreases endogenous opioid produc-tion; depresses gonadotroph respon-siveness to GnRH; stimulates GH gene expression by the pituitary; inhibits GH secretion via the hypothalamus	Scanty menses due to suppressed gonadotroph sensitivity to GnRH; suppressed GH secretion by hypo-thalamic action; minimal suppres-sion of the TRH-TSH axis	Scanty menses by upregulated CRH-endogenous opioid pathway-mediated suppression of GnRH; suppressed GH secre-tion; hypothyroidism (if present) is due to autoimmune mechanism
	Pancreas	Inhibits insulin secretion by decreasing the efficacy of cytoplasmic Ca^{2+} on the exocytotic process	Absolute hyperinsulinemia with relative hypoinsulinemia (lower plasma insulin than expected for the degree of hyperglycemia)	Absolute hypoinsu-linemia with relative hyperinsulinemia
	Adrenal medulla	Increases PNMT expression and activity (epinephrine synthesis)	Increased responses to sympathoad-renal activation	Decreased responses to sympathoadrenal activation
	Carrier proteins (CBG, SHBG, TBG)	Decreases all major hormone-binding proteins	Decrease in total T_4, free T_4 remains normal	
Immune system	Thymus, lymphocytes	Causes age-related involution of the thymus; induces thymic atrophy	Immunocompromised state; lymphocytopenia	Relative lymphocytosis in peripheral blood
	Monocytes	Inhibits monocyte proliferation and antigen presentation; decreased production of IL-1, IL-6, and TNFα	Monocytopenia in peripheral blood	Monocytosis in periph-eral blood
	Granulocytes	Demargination of neutrophils by suppressing the expression of adhesion molecules	Peripheral blood: granulocytosis, eosinopenia	Peripheral blood: granulocytopenia, eosinophilia

(Continued)

TABLE 9–2 The main targets and actions of glucocorticoids and the consequences of Cushing disease and Addison disease. (*Continued*)

Target System	Specific Target	Physiologic Function	Cushing Disease	Addison Disease
	Inflammatory response	Inhibition of inflammation by inhibiting PLA_2, thereby inhibiting production of leukotrienes and prostaglandins; suppresses COX-2 expression		
	Erythrocytes	No significant effect	Increased hemoglobin and hematocrit are due to ACTH-mediated overproduction of androgens	Anemia is more pronounced in women and is due to loss of adrenal androgens; anemia may be related to direct autoimmune targeting of gastric parietal cells
Skin and connective tissue		Antiproliferative for fibroblasts and keratinocytes	Easy bruisability due to dermal atrophy; striae or sites of increased tension, especially sites of adipose tissue accumulation; poor wound healing; hirsutism and acne are due to ACTH-mediated increase of adrenal androgens; hyperpigmentation is a direct effect of ACTH on melanocortin 1 receptors	Darkening of the skin is due to ACTH-mediated stimulation of epidermal melanocortin 1 receptors; vitiligo may occur due to direct autoimmune destruction of melanocytes in circumscribed areas
Breast	Mammary epithelium	Mandatory requirement for lactation	Cushing disease may be associated with galactorrhea	Addison disease is not associated with galactorrhea
Lung	Type 2 alveolar cell	Stimulation of surfactant production		
Cardiovascular system	Heart	Increased contractility	Hypertension	Lower peripheral resistance; hypotension with further postural decrease in blood pressure (orthostatic hypotension); low-voltage ECG
	Vasculature	Increased vascular reactivity to vasoconstrictors (catecholamines, angiotensin II)		
Na^+, K^+, and ECF volume	Kidney	Increased GFR and nonphysiologic actions on mineralocorticoid receptors	Hypokalemic alkalosis, increased ECF volume due to mineralocorticoid activity (increased DOC, saturation of type 2 11β-hydroxysteroid dehydrogenase by high levels of cortisol)	Hyponatremia, hyperkalemic acidosis, and decreased ECF volume are mainly due to loss of mineralocorticoid activity
	Posterior pituitary		Hyponatremia due to SIADH	Increased ADH mainly via hypovolemia-related baroreceptor mechanism
Psychiatric parameters of CNS function	Mood	Eucortisolemia maintains emotional balance	Initially, euphoria; long-term, depression, psychosis	Depression
	Appetite	Increases appetite	Hyperphagia	Decreased appetite in spite of improved taste and smell
	Sleep	Suppression of REM sleep	Sleep disturbances	
	Memory	Sensitizes hippocampal glutamate receptors, induces atrophy of dendrites	Impaired memory, bilateral hippocampal atrophy	
	Eye	Increasing intraocular pressure	Cataract formation; increased intraocular pressure	Decreased intraocular pressure

Modified with permission from Kacsoh B. *Endocrine Physiology*. New York: McGraw-Hill Education; 2000.

present in equimolar concentrations. Corticosterone and aldosterone have substantial affinity for the glucocorticoid receptor; however, their plasma concentrations are normally much lower than that of cortisol, and thus these steroids do not have significant physiologic glucocorticoid effects.

B. Antagonists   Glucocorticoid antagonists bind to the glucocorticoid receptors but do not elicit the nuclear events required to cause a glucocorticoid response. These steroids compete with agonist steroids such as cortisol for the receptors and thus inhibit agonist responses. Other steroids have partial agonist activity when present alone (ie, they elicit a partial glucocorticoid response). However, in sufficient concentration, they compete with agonist steroids for the receptors and thus competitively inhibit agonist responses; that is, these partial agonists may function as partial antagonists in the presence of more active glucocorticoids. Steroids such as progesterone, 11-deoxycortisol, DOC, testosterone, and 17β-estradiol have antagonist or partial agonist-partial antagonist effects; however, the physiologic role of these hormones in glucocorticoid action is probably negligible, because they circulate in low concentrations. The antiprogestational agent mifepristone has substantial glucocorticoid antagonist properties and has been used to block glucocorticoid action in patients with Cushing syndrome.

Intermediary Metabolism (Table 9–2)

Glucocorticoids in general inhibit DNA synthesis. In addition, in most tissues they inhibit RNA and protein synthesis and accelerate protein catabolism. These actions provide substrate for intermediary metabolism; however, accelerated catabolism also accounts for the deleterious effects of glucocorticoids on muscle, bone, connective tissue, and lymphatic tissues. In contrast, RNA and protein synthesis in liver is stimulated.

A. Hepatic glucose metabolism   Glucocorticoids increase hepatic gluconeogenesis by stimulating the gluconeogenic enzymes phosphoenolpyruvate carboxykinase and glucose-6-phosphatase. They have a permissive effect in that they increase hepatic responsiveness to the gluconeogenic hormone glucagon, and they also increase the release of substrates for gluconeogenesis from peripheral tissues, particularly muscle. This latter effect may be enhanced by the glucocorticoid-induced reduction in peripheral amino acid uptake and protein synthesis. Glucocorticoids also increase glycerol and free fatty acid release by lipolysis and increase muscle lactate release. They enhance hepatic glycogen synthesis and storage by stimulating glycogen synthetase activity and to a lesser extent by inhibiting glycogen breakdown. These effects are insulin dependent.

B. Peripheral glucose metabolism   Glucocorticoids also alter carbohydrate metabolism by inhibiting peripheral glucose uptake in muscle and adipose tissue. This effect and the others described earlier may result in increased insulin secretion in states of chronic glucocorticoid excess.

C. Effects on adipose tissue   In adipose tissue, the predominant effect is increased lipolysis with release of glycerol and free fatty acids. This is partially due to direct stimulation of lipolysis by glucocorticoids, but it is also contributed to by decreased glucose uptake and enhancement by glucocorticoids of the effects of lipolytic hormones. Although glucocorticoids are lipolytic, increased fat deposition is a classic manifestation of glucocorticoid excess. This paradox may be explained by the increased appetite caused by high levels of these steroids and by the lipogenic effects of the hyperinsulinemia that occurs in this state. The reason for abnormal fat deposition and distribution in states of cortisol excess is unknown. In these instances, fat is classically deposited centrally in the face, cervical area, trunk, and abdomen; the extremities are usually spared.

D. Summary   The effects of the glucocorticoids on intermediary metabolism can be summarized as follows:

1. Effects are minimal in the fed state. However, during fasting, glucocorticoids contribute to the maintenance of plasma glucose levels by increasing gluconeogenesis and decreasing uptake of glucose by adipose tissue.
2. Hepatic glucose production is enhanced, as is hepatic RNA and protein synthesis.
3. The effects on muscle are catabolic (ie, decreased glucose uptake and metabolism, decreased protein synthesis, and increased release of amino acids). This provides amino acid substrates for gluconeogenesis in the liver.
4. In adipose tissue, lipolysis is stimulated. This increases fatty acid delivery to the liver where their metabolism provides energy to support gluconeogenesis.
5. In glucocorticoid deficiency, hypoglycemia may result, whereas in states of glucocorticoid excess there may be hyperglycemia, hyperinsulinemia, muscle wasting, and weight gain with abnormal fat distribution.

Effects on Other Tissues and Functions (Table 9–2)

A. Connective tissue   Glucocorticoids in excess inhibit fibroblasts, lead to loss of collagen and connective tissue, and thus result in thinning of the skin, easy bruising, stria formation, and poor wound healing.

B. Bone   The physiologic role of glucocorticoids in bone metabolism and calcium homeostasis is unknown; however, in excess, they have major deleterious effects. Glucocorticoids directly inhibit bone formation by decreasing osteoblast proliferation and the synthesis of RNA, protein, collagen, and hyaluronate. Initially, supraphysiologic doses of glucocorticoids also stimulate bone resorption, at least in part via activation of the receptor activator of nuclear factor kappa B (RANK)-ligand/ RANK signaling that is osteoclastogenic. This leads to osteolysis and increased biochemical markers of bone turnover. In addition, glucocorticoids may potentiate the proresorptive actions of parathyroid hormone (PTH) and 1,25-dihydroxycholecalciferol

$[1,25(OH)_2D_3]$ on bone, and this may further contribute to net bone resorption. Chronically, however, the key effect to promote bone loss is the detrimental impact on bone formation (see Chapter 8).

C. Calcium metabolism   Glucocorticoids also have other major effects on mineral homeostasis. They markedly reduce intestinal calcium absorption, which tends to lower serum calcium. This theoretically promotes a state of secondary hyperparathyroidism to maintain the serum calcium within the normal range. In reality, however, chronic elevations in intact PTH levels have been difficult to demonstrate consistently in patients on glucocorticoid therapy. The mechanism of decreased intestinal calcium absorption is thought to result from antagonism of $1,25(OH)_2D$ action in the intestine. It is not due to decreased synthesis or decreased serum levels of the active vitamin D metabolites; in fact, $1,25(OH)_2D$ levels are normal or even increased in the presence of glucocorticoid excess. Increased $1,25(OH)_2D$ synthesis in this setting may result from decreased serum phosphorus levels (see later), increased PTH secretion, direct stimulation by glucocorticoids of renal 1α-hydroxylase, and or reduced target cell (intestine) responsiveness. Glucocorticoids increase urinary calcium excretion, and hypercalciuria is a frequent feature of cortisol excess. They also reduce the tubular reabsorption of phosphate, leading to phosphaturia and decreased serum phosphorus concentrations.

Thus, glucocorticoids in excess result in negative calcium balance, with decreased calcium absorption and increased urinary calcium excretion. Serum calcium levels are maintained, but at the expense of net bone resorption. Decreased bone formation and increased resorption ultimately result in the disabling osteoporosis that is often a major complication of spontaneous and iatrogenic glucocorticoid excess (see Chapter 8).

D. Growth and development   Glucocorticoids accelerate the development of a number of systems and organs in fetal and differentiating tissues, although the mechanisms are unclear. As discussed earlier, glucocorticoids are generally inhibitory, and these stimulatory effects may be due to glucocorticoid interactions with other growth factors. Examples of these development-promoting effects are increased surfactant production in the fetal lung and the accelerated development of hepatic and gastrointestinal enzyme systems.

Glucocorticoids in excess inhibit growth in children, and this adverse effect is a major complication of therapy. This may be a direct effect on bone cells, although decreased growth hormone (GH) secretion and insulin-like growth factor I generation also contribute (see Chapter 6).

E. Blood cells and immunologic function

1. **Erythrocytes**—Glucocorticoids have little effect on erythropoiesis and hemoglobin concentration. Although mild polycythemia and anemia may be seen in Cushing syndrome and Addison disease, respectively, these alterations are more likely to be secondary to altered androgen metabolism.

2. **Leukocytes**—Glucocorticoids influence both leukocyte movement and function. Thus, glucocorticoid administration increases the number of intravascular polymorphonuclear neutrophils/leukocytes (PMNs) by increasing PMN release from bone marrow, by increasing the circulating half-life of PMNs, and by decreasing PMN movement out of the vascular compartment. Glucocorticoid administration reduces the number of circulating lymphocytes, monocytes, and eosinophils, mainly by increasing their movement out of the circulation and may be lympholytic as well. The converse (ie, neutropenia, lymphocytosis, monocytosis, and eosinophilia) is seen in adrenal insufficiency. Glucocorticoids also decrease the migration of inflammatory cells (PMNs, monocytes, and lymphocytes) to sites of injury, and this is probably a major mechanism of the anti-inflammatory actions and increased susceptibility to infection that occur following chronic administration. Glucocorticoids also decrease lymphocyte production and the mediator and effector functions of these cells.

3. **Immunologic effects**—Glucocorticoids influence multiple aspects of immunologic and inflammatory responsiveness, including the mobilization and function of leukocytes, as discussed earlier. They inhibit phospholipase A_2, a key enzyme in the synthesis of prostaglandins. This inhibition is mediated by a class of peptides called lipocortins or annexins. They also impair the release of effector substances such as the lymphokine IL-1, antigen processing, antibody production and clearance, and other specific bone marrow-derived and thymus-derived lymphocyte functions. The immune system, in turn, affects the HPA axis; IL-1 stimulates the secretion of CRH and ACTH. Although traditionally used as anti-inflammatory and/or immunosuppressive agents, glucocorticoids, especially at lower doses, also have stimulatory and permissive effects on the inflammatory response to injury.

F. Cardiovascular function   Glucocorticoids may increase cardiac output, and they also increase peripheral vascular tone, possibly by augmenting the effects of other vasoconstrictors (eg, the catecholamines). Glucocorticoids also regulate expression of adrenergic receptors. Thus, refractory shock may occur when the glucocorticoid-deficient individual is subjected to stress. Glucocorticoids in excess may cause hypertension independently of their mineralocorticoid effects. Although the incidence and the precise cause of this problem are unclear, it is likely that the mechanism involves the renin-angiotensin system; glucocorticoids regulate renin substrate, angiotensinogen, the precursor of angiotensin I.

G. Renal function   These steroids affect water and electrolyte balance by actions mediated either by mineralocorticoid receptors (sodium retention, hypokalemia, and hypertension) or via glucocorticoid receptors (increased glomerular filtration rate due to increased cardiac output or due to direct renal effects on salt and water retention). Thus, corticosteroids such as betamethasone or dexamethasone, which have little mineralocorticoid activity, increase sodium and water excretion. Glucocorticoid-deficient subjects have decreased glomerular filtration rates and are unable to excrete a water load. This may be further aggravated by increased ADH secretion, which may occur in glucocorticoid deficiency.

H. Central nervous system function Glucocorticoids readily enter the brain, and although their physiologic role in central nervous system function is unknown, their excess or deficiency may profoundly alter behavior and cognitive function.

1. **Excessive glucocorticoids**—In excess, the glucocorticoids initially cause euphoria; however, with prolonged exposure, a variety of psychologic abnormalities occur, including irritability, emotional lability, and depression. Hyperkinetic or manic behavior is less common; overt psychoses occur in a small number of patients, particularly those with underlying bipolar disorder. Many patients also note impairment in cognitive functions, most commonly memory and concentration. Other central effects include increased appetite, decreased libido, and insomnia, with decreased rapid eye movement sleep and increased stage II sleep.
2. **Decreased glucocorticoids**—Patients with Addison disease are apathetic and depressed and tend to be irritable, negativistic, and reclusive. They have decreased appetite but increased sensitivity of taste and smell mechanisms.

I. Effects on other hormones

1. **Thyroid function**—Glucocorticoids in excess affect thyroid function. Thyroid-stimulating hormone (TSH) synthesis and release are inhibited by glucocorticoids, and TSH responsiveness to thyrotropin-releasing hormone (TRH) is frequently subnormal. Although basal levels are usually normal, the TSH concentration may be acutely low in patients treated with moderate- to high-dose glucocorticoids. Serum total thyroxine (T_4) concentrations are usually low normal because of a decrease in thyroxine-binding globulin, but free T_4 levels are normal. Total and free T_3 (triiodothyronine) concentrations may be low, because glucocorticoid excess decreases the conversion of T_4 to T_3 and increases conversion to reverse T_3. Despite these alterations, manifestations of hypothyroidism are not apparent.
2. **Gonadal function**—Glucocorticoids also affect gonadotropin and gonadal function. In males, they inhibit gonadotropin secretion, as evidenced by decreased responsiveness to administered gonadotropin-releasing hormone (GnRH) and subnormal plasma testosterone concentrations. In females, glucocorticoids also suppress luteinizing hormone (LH) responsiveness to GnRH, resulting in suppression of estrogens and progestins with inhibition of ovulation and amenorrhea.

J. Miscellaneous effects

1. **Peptic ulcer**—The role of steroid excess in the production or reactivation of peptic ulcer disease is controversial. However, there appears to be a modest independent effect of glucocorticoids to promote peptic ulcer disease (relative risk about 1.4), and when this effect is combined with that of nonsteroidal anti-inflammatory drugs, there is a synergistic interaction that considerably increases the risk.
2. **Ophthalmologic effects**—Intraocular pressure varies with the level of circulating glucocorticoids and parallels the circadian variation of plasma cortisol levels. In addition, glucocorticoids in excess increase intraocular pressure in patients with open-angle glaucoma. Glucocorticoid therapy may also cause cataract formation. Central serous chorioretinopathy, an accumulation of subretinal detachment, may also complicate endogenous or exogenous glucocorticoid excess.

ADRENAL ANDROGENS

The direct biologic activity of the adrenal androgens (androstenedione, DHEA, and DHEA sulfate) is minimal, and they function primarily as precursors for peripheral conversion to the active androgenic hormones, testosterone and dihydrotestosterone. Thus, DHEA sulfate secreted from the adrenal undergoes limited conversion to DHEA; this peripherally converted DHEA and that secreted by the adrenal cortex can be further converted in peripheral tissues to androstenedione, the immediate precursor of the active androgens.

The actions of testosterone and dihydrotestosterone are described in Chapter 12. This section deals only with the adrenal contribution to androgenicity.

Effects in Males

In males with normal gonadal function, the conversion of adrenal androstenedione to testosterone accounts for less than 5% of the production rate of this hormone, and thus the physiologic effect is negligible. In adult males, excessive adrenal androgen secretion has no clinical consequences; however, in boys, it causes premature penile enlargement and early development of secondary sexual characteristics.

Effects in Females

In females, the adrenal substantially contributes to total androgen production by the peripheral conversion of androstenedione to testosterone. In the follicular phase of the menstrual cycle, adrenal precursors account for two-thirds of testosterone production and one-half of dihydrotestosterone production. During midcycle, the ovarian contribution increases, and the adrenal precursors account for only 40% of testosterone production.

In females, abnormal adrenal function as seen in Cushing syndrome, adrenal carcinoma, and congenital adrenal hyperplasia results in excessive secretion of adrenal androgens, and their peripheral conversion to testosterone results in androgen excess, manifested by acne, hirsutism, and virilization.

LABORATORY EVALUATION

Cortisol and the adrenal androgens are measured by specific plasma assays. Certain urinary assays, particularly measurement of 24-hour urine free cortisol, are also useful. In addition, plasma concentrations of ACTH can be determined. The assays for plasma steroids commonly used measure the total hormone concentrations and are therefore influenced by alterations in plasma-binding proteins. Furthermore, because ACTH and the plasma concentrations of the adrenal hormones fluctuate markedly (see Figure 9–6), single plasma measurements are frequently unreliable. Thus, plasma levels must be interpreted cautiously, and more

specific diagnostic information is usually obtained by performing appropriate dynamic tests (stimulation and suppression) or other tests that reflect cortisol secretory rate.

Plasma ACTH

A. Methods of measurement   Plasma ACTH measurements are extremely useful in the diagnosis of pituitary-adrenal dysfunction. The typical reference range for plasma ACTH, using a sensitive immunoradiometric assay (IRMA) or immunochemiluminometric assay (ICMA), is 9 to 52 pg/mL (2-11.1 pmol/L).

B. Interpretation   Plasma ACTH levels are most useful in differentiating pituitary causes from adrenal causes of adrenal dysfunction as follows:

1. In adrenal insufficiency due to primary adrenal disease, plasma ACTH levels are elevated. Conversely, in pituitary ACTH deficiency and secondary hypoadrenalism, plasma ACTH levels are frankly low or inappropriately within the normal reference range.
2. In Cushing syndrome due to primary glucocorticoid-secreting adrenal tumors, plasma ACTH is suppressed, and typically less than the lower limit of the normal reference range. In patients with Cushing disease (pituitary ACTH hypersecretion), plasma ACTH levels are inappropriately normal or elevated.
3. Plasma ACTH levels are usually markedly elevated in the ectopic ACTH syndrome, but there is a considerable amount of overlap with levels seen in Cushing disease. In addition, values lower than expected may be observed rarely in ectopic ACTH syndrome when the two-site IRMA is used; this assay may not detect high-molecular-weight precursors of ACTH that can be produced in ectopic ACTH syndrome.
4. Plasma ACTH levels are also elevated in patients with the common forms of congenital adrenal hyperplasia and are useful in the diagnosis and management of these disorders (see Chapter 14).

Plasma Cortisol

A. Methods of measurement   The most common methods of measurement of plasma cortisol are radioimmunoassay, enzyme-linked immunosorbent assay (ELISA), high-performance liquid chromatography (HPLC), and liquid chromatography tandem mass spectroscopy (LC/MS/MS). These methods measure total cortisol (both bound and free) in plasma. Most commonly used drugs and medications do not interfere with this assay. In contrast to radioimmunoassays, HPLC and LC/MS/MS do not demonstrate cross-reactivity with synthetic glucocorticoids.

B. Interpretation   The diagnostic utility of single plasma cortisol concentrations is limited by the episodic nature of cortisol secretion and its appropriate elevations during stress. As explained later, more information is obtained by dynamic testing of the HPA axis. Reference ranges for cortisol depend on the methodology used. With the introduction of more specific cortisol assays (eg, LC/MS/MS), the normal range for cortisol in the basal state and after dynamic testing may be lower than currently accepted ranges.

1. **Normal values**—Normal plasma cortisol levels vary with the method used and time of day the sample is obtained. With radioimmunoassay, levels at 8 AM range from 3 to 20 μg/dL (80-550 nmol/L) and average 10 to 12 μg/dL (275.9-331.1 nmol/L). Values obtained later in the day are lower and at 4 PM are approximately half of morning values. At 10 PM to 2 AM, the plasma cortisol concentrations by these methods are usually less than 3 μg/dL (80 nmol/L).
2. **Levels during stress**—Cortisol secretion increases in patients who are acutely ill, during surgery, and following trauma when plasma concentrations may reach greater than 40 to 60 μg/dL (1100-1655 nmol/L).
3. **High-estrogen states**—The total plasma cortisol concentration is also elevated with increased CBG-binding capacity, which occurs most commonly when circulating estrogen levels are high (eg, during pregnancy and when exogenous estrogens or oral contraceptives are being used). In these situations, plasma cortisol may reach levels two to three times normal.
4. **Other conditions**—CBG levels may be increased or decreased in other situations, as discussed earlier in the sections on circulation and metabolism. Total plasma cortisol concentrations may also be increased in severe anxiety, endogenous depression, starvation, anorexia nervosa, alcoholism, and chronic kidney disease.

Salivary Cortisol

Cortisol in the saliva is in equilibrium with the free and biologically active cortisol in the blood. Salivary cortisol concentrations are not affected by changes in serum cortisol–binding proteins, by salivary flow or composition, and they are stable at room temperature for many days. Measurements of salivary cortisol can be obtained from late-night, ambulatory saliva samples, which are used as a means of establishing the presence or absence of Cushing syndrome. Salivary cortisol may also be used to obtain accurate free cortisol levels in patients with abnormal serum-binding proteins.

A. Methods of measurement   Saliva can easily be sampled at home by the patient using a variety of techniques, including the use of a commercially available sampling device or by passive drooling. The same methods that are employed to measure plasma cortisol, radioimmunoassay, ELISA, HPLC, and LC/MS/MS, are used to measure salivary cortisol concentrations.

B. Interpretation

1. **Normal values**—Reference ranges for late-night salivary cortisol concentrations are dependent on the assays employed, but with radioimmunoassay and ELISA, the normal salivary cortisol level at midnight is generally less than 0.15 μg/dL (4 nmol/L).
2. **Diagnostic utility**—As with single plasma cortisol measurements, randomly obtained salivary cortisol concentrations are of limited use. However, plasma and salivary cortisol in normal individuals reach a nadir from 10 PM to 2 AM. Patients with Cushing syndrome do not reach a normal nadir at this time, and several studies have shown that elevated late-night salivary cortisol is a sensitive and specific diagnostic test for Cushing syndrome.

Plasma Free Cortisol

This assay measures the free or biologically active fraction of plasma cortisol and, thus, is not influenced by alterations in CBG and serum albumin. The best current method involves equilibrium dialysis to separate bound and free hormone followed by measurement of the free fraction by radioimmunoassay. In patients with normal serum albumin and CBG, the free cortisol correlates well with total cortisol levels; however, in patients with disturbances in plasma-binding proteins, this correlation is lost.

Although this assay is not in common use, it is of importance in the assessment of adrenal function in patients with critical illness. In these patients the measurement of total plasma cortisol and its response to ACTH has overestimated the number of patients with adrenal insufficiency, especially in the most severely ill patients who frequently have hypoalbuminemia and subnormal CBG levels.

Urinary Corticosteroids

A. Free cortisol

1. **Methods of measurement**—The assay of unbound cortisol excreted in the urine is another method for the diagnosis of Cushing syndrome. Normally, less than 1% of the secreted cortisol is excreted unchanged in the urine. However, in states of excess secretion, the binding capacity of CBG is exceeded, and plasma free cortisol increases, as does its urinary excretion. Urine free cortisol is measured in a 24-hour urine collection by HPLC, radioimmunoassay, and now, most commonly, by LC/MS/MS.
2. **Normal values**—HPLC and LC/MS/MS provide the most specific measurement of cortisol and are the current procedures of choice. The reference range for urine free cortisol assayed by HPLC or LC/MS/MS is generally 5 to 50 μg/24 h (14-135 nmol/24 h).
3. **Diagnostic utility**—This method has historically been the gold standard for diagnosis of Cushing syndrome, but recent studies questions the sensitivity of this test. Interestingly, the levels of urine free cortisol may be elevated in the same conditions that increase total plasma cortisol (see earlier), including a slight elevation during pregnancy and women taking oral contraceptives. This test is not useful in adrenal insufficiency, because it lacks sensitivity at low levels and because low cortisol excretion is often found in normal persons.

B. 17-Hydroxycorticosteroids
These urinary steroids are largely of historical interest and should not be measured at present because of the greater utility of plasma cortisol and urine free cortisol measurements.

Dexamethasone Suppression Tests

A. Low-Dose test
This procedure is used to establish the presence of Cushing syndrome regardless of its cause. Dexamethasone, a potent glucocorticoid, normally suppresses pituitary ACTH release with a resulting fall in plasma and urine cortisol, thus assessing feedback inhibition of the HPA axis. In Cushing syndrome, this mechanism is abnormal, and steroid secretion fails to be suppressed in the normal way.

Dexamethasone, in the doses used, does not interfere with the measurement of plasma and urinary cortisol.

The overnight 1-mg dexamethasone suppression test is commonly used as a screening test for Cushing syndrome. Dexamethasone, 1 mg orally, is given as a single dose at 11:00 PM, and the following morning a plasma sample is obtained for cortisol determination. Cushing syndrome is probably excluded if the serum or plasma cortisol level is less than 1.8 μg/dL (50 nmol/L). Utilizing this criterion leads to very high sensitivity but also has an increased false-positive rate and results in a specificity of 80% to 90%.

False-positive results occur in hospitalized and chronically ill patients. Acute illness, depression, anxiety, alcoholism, high-estrogen states, and uremia may also cause false-positive results. Patients taking phenytoin, barbiturates, and other inducers of hepatic microsomal enzymes may have accelerated metabolism of dexamethasone and thus fail to achieve adequate plasma levels to suppress ACTH.

B. High-dose tests
High-dose dexamethasone suppression testing has historically been used to differentiate Cushing disease (pituitary ACTH hypersecretion) from ectopic ACTH and adrenal tumors. This rationale has been based on the fact that in some patients with Cushing disease, the HPA axis is suppressible with supraphysiologic doses of glucocorticoids, whereas cortisol secretion is autonomous in patients with adrenal tumors and in most patients with the ectopic ACTH syndrome. Unfortunately, exceptions to these rules are so common that high-dose dexamethasone suppression testing must be interpreted with extreme caution.

1. **Overnight high-dose dexamethasone suppression test**—After a baseline morning cortisol specimen is obtained, a single dose of dexamethasone, 8 mg orally, is administered at 11:00 PM and plasma cortisol is measured at 8:00 AM the following morning. Generally, patients with Cushing disease suppress plasma cortisol level to less than 50% of baseline values—in contrast to patients with the ectopic ACTH syndrome, who fail to suppress to this level. Patients with cortisol-producing adrenal tumors also fail to suppress: their cortisol secretion is autonomous, and ACTH secretion is already suppressed by the high endogenous levels of cortisol.
2. **Two-day high-dose dexamethasone suppression test**—This test has been used to differentiate patients with Cushing disease from those with an ectopic ACTH syndrome. It is performed by administering dexamethasone, 2 mg orally every 6 hours for 2 days. Twenty-four-hour urine samples are collected before and on the second day of dexamethasone administration. Patients with Cushing disease have a reduction of urine cortisol excretion to less than 50% of baseline values, whereas those with adrenal tumors or the ectopic ACTH syndrome usually have little or no reduction in urinary cortisol excretion. However, some patients with an ectopic ACTH-secreting neoplasm suppress steroid secretion with high doses of dexamethasone, and some patients with pituitary ACTH-dependent Cushing syndrome fail to suppress to these levels. The diagnostic sensitivity, specificity, and accuracy of the high-dose dexamethasone suppression test are only about 80%. Better specificity and accuracy can be achieved by utilizing different criteria. However, it has become increasingly clear that high-dose

dexamethasone suppression testing, regardless of the criteria employed, cannot distinguish pituitary from nonpituitary ACTH hypersecretion with certainty. Currently, the only biochemical test with a high enough diagnostic accuracy to distinguish between these causes is inferior petrosal sinus ACTH sampling.

Pituitary-Adrenal Reserve

These tests are used to evaluate the ability of the HPA axis to respond to stress. ACTH administration directly stimulates adrenal secretion; metyrapone inhibits cortisol synthesis, thereby stimulating pituitary ACTH secretion; and insulin-induced hypoglycemia stimulates ACTH release by increasing CRH secretion. More recently, CRH has been utilized to directly stimulate pituitary corticotrophs to release ACTH. The relative utility of these procedures is discussed later in the section on adrenocortical insufficiency and also in Chapter 4.

A. ACTH stimulation testing

1. **Procedure and normal values—High-dose ACTH stimulation test**. The rapid ACTH stimulation test measures the acute adrenal response to ACTH and is used to diagnose both primary and secondary adrenal insufficiency. A synthetic human α^{1-24}-ACTH called tetracosactin or cosyntropin is used. Fasting is not required, and the test may be performed at any time of the day. A baseline cortisol sample is obtained; cosyntropin is administered in a dose of 250 µg intramuscularly or intravenously; and additional samples for plasma cortisol are obtained at 30 and/or 60 minutes following the injection. Because the peak concentration of ACTH with this test achieves a pharmacologic level exceeding 10,000 pg/mL, this study assesses maximal adrenocortical capacity. The peak cortisol response, 30 to 60 minutes later, should exceed 18 to 20 µg/dL (>497-552 nmol/L). The 30-minute peak cortisol response to ACTH is constant and is unrelated to the basal cortisol level. In fact, there is no difference in the peak cortisol level at 30 minutes regardless of whether 250 µg, 5 µg, or even 1 µg of ACTH is administered. Early studies of the salivary or plasma free cortisol response to ACTH suggest that this method may be useful in patients with abnormal levels of cortisol-binding proteins.

    Low-dose ACTH stimulation test. Since the standard or high-dose test may be normal in patients with partial secondary adrenal insufficiency, a low-dose (1 µg) ACTH stimulation test was developed. Although some studies reported increased sensitivity and specificity, two meta-analyses did not establish a distinct advantage for the low-dose test.

    To perform the low-dose test appropriately it must be undertaken in the morning and the ACTH must be given intravenously. In addition, technical limitations are significant (ie, ACTH is available only in 250-µg vials, it is unstable in solution, and adheres to glass and plastic syringes and IV tubing). Thus, there are serious issues as to the accuracy of the dose of ACTH during reconstitution, dilution, and IV injection. For these reasons, the low-dose ACTH stimulation test is not practical for routine clinical practice.

2. **Subnormal responses**—If the cortisol response to the rapid ACTH stimulation test is inadequate, adrenal insufficiency is present. In primary adrenal insufficiency, destruction of cortical cells reduces cortisol secretion and increases pituitary ACTH secretion. Therefore, the adrenal is already maximally stimulated, and there is no further increase in cortisol secretion when exogenous ACTH is given; that is, there is decreased adrenal reserve. In secondary adrenal insufficiency due to ACTH deficiency, there is atrophy of the zonae fasciculata and reticularis, and the adrenal thus is either hyporesponsive or unresponsive to acute stimulation with exogenous ACTH. In either primary or secondary types, a subnormal response to the rapid ACTH stimulation test accurately predicts deficient responsiveness of the axis to insulin hypoglycemia, metyrapone, and surgical stress.

3. **Normal responses**—A normal response to the rapid ACTH stimulation test excludes both primary adrenal insufficiency (by directly assessing adrenal reserve) and overt secondary adrenal insufficiency with adrenal atrophy. However, a normal response does not rule out partial ACTH deficiency (decreased pituitary reserve) in patients whose basal ACTH secretion is sufficient to prevent adrenocortical atrophy and patients with recently developed secondary adrenal insufficiency who have not yet undergone adrenal atrophy. These patients may be unable to further increase ACTH secretion and thus may have subnormal pituitary ACTH responsiveness to stress or hypoglycemia. In such patients, further testing with metyrapone, hypoglycemia, or CRH may be indicated. For further discussion, see the section on Diagnosis of Adrenocortical Insufficiency.

B. Metyrapone testing

Metyrapone testing has been used to diagnose adrenal insufficiency and to assess pituitary-adrenal reserve. The test procedures are detailed in Chapter 4. Metyrapone blocks cortisol synthesis by inhibiting the 11β-hydroxylase enzyme that converts 11-deoxycortisol to cortisol. This stimulates ACTH secretion, which in turn increases the secretion and plasma levels of 11-deoxycortisol. The overnight metyrapone test is most commonly used and is best suited to patients with suspected pituitary ACTH deficiency; patients with suspected primary adrenal failure are usually evaluated with the rapid ACTH stimulation test as described earlier and discussed in the section on Diagnosis of Adrenocortical Insufficiency. The normal response to the overnight metyrapone test is a plasma 11-deoxycortisol level greater than 7 mcg/dL (200 nmol/L) and a plasma ACTH level greater than 100 pg/mL (22 pmol/L) and indicates both normal ACTH secretion and adrenal function. A subnormal response establishes adrenocortical insufficiency. A normal response to metyrapone accurately predicts normal stress responsiveness of the hypothalamic-pituitary axis and correlates well with responsiveness to insulin-induced hypoglycemia. Metyrapone is only available directly from the manufacturer.

C. Insulin-induced hypoglycemia testing

The details of this procedure are described in Chapter 4. Hypoglycemia induces a central nervous system stress response, increases CRH release, and in this way increases ACTH and cortisol secretion. It therefore measures the integrity of the axis and its ability to respond to stress. In the non-acutely ill individual, the normal plasma cortisol response is an increment greater than 8 µg/dL (220 nmol/L) and a peak level greater than 18 to 20 µg/dL

(497-552 nmol/L). The plasma ACTH response to hypoglycemia is usually greater than 100 pg/mL (22 pmol/L). A normal plasma cortisol response to hypoglycemia excludes adrenal insufficiency and decreased pituitary reserve. Thus, patients with normal responses do not require cortisol therapy during illness or surgery.

D. CRH testing The procedure for CRH testing is described in Chapter 4. ACTH responses are exaggerated in patients with primary adrenal failure and absent in patients with hypopituitarism. Delayed responses may occur in patients with hypothalamic disorders. CRH testing has also been used to differentiate among the causes of Cushing syndrome (see later).

Androgens

Androgen excess is usually evaluated by the measurement of basal levels of these hormones, because suppression and stimulation tests are not as useful as in disorders affecting glucocorticoids. Assays are available for total plasma levels of DHEA, DHEA sulfate, androstenedione, testosterone, and dihydrotestosterone. Plasma-free testosterone (ie, testosterone not bound to SHBG) can be measured and is a more direct measure of circulating biologically active testosterone than the total plasma level, but commercially available methods vary widely in their accuracy (see Chapter 12).

DISORDERS OF ADRENOCORTICAL INSUFFICIENCY

Deficient adrenal production of glucocorticoids or mineralocorticoids results in adrenocortical insufficiency, which is either the consequence of destruction or dysfunction of the cortex (primary adrenocortical insufficiency, or Addison disease) or secondary to deficient pituitary ACTH secretion (secondary adrenocortical insufficiency). Glucocorticoid therapy is the most common cause of secondary adrenocortical insufficiency.

PRIMARY ADRENOCORTICAL INSUFFICIENCY (ADDISON DISEASE)

Etiology and Pathology (Figure 9–8)

The etiology of primary adrenocortical insufficiency has changed over time. Prior to 1920, tuberculosis was the major cause of adrenocortical insufficiency. Since 1950, autoimmune adrenalitis with adrenal atrophy has accounted for about 80% of cases. It is associated with a high incidence of other immunologic and autoimmune endocrine disorders (see later). Causes of primary adrenal insufficiency are listed in Table 9–3. Primary adrenocortical insufficiency, or Addison disease, is uncommon, with a reported prevalence of 35 to 140 per million population. When part of the polyglandular system, it is more common in females, with a

female:male ratio of approximately 2:1. Addison disease is usually diagnosed in the third to fifth decades.

A. Autoimmune adrenocortical insufficiency Pathologically, the adrenals are small and atrophic, and the capsule is thickened. The adrenal medulla is preserved, although cortical cells are largely absent, show degenerative changes, and are surrounded by a fibrous stroma and the characteristic lymphocytic infiltrates.

Autoimmune Addison disease is frequently accompanied by other immune disorders. There are two different syndromes in which autoimmune adrenal insufficiency may occur. The best characterized one is known as autoimmune polyendocrine syndrome type 1 (APS-1). This is generally an autosomal recessive disorder that usually presents in childhood and is accompanied by hypoparathyroidism, adrenal failure, and mucocutaneous candidiasis. APS-1 results in most cases from a mutation of the autoimmune regulator gene (*AIRE*), which is located on chromosome 21q22.3. These patients have impaired deletion of auto-reactive T cells during T cell development in the thymus. APS-1 has no relationship to human leukocyte antigen (HLA) and is often associated with autoimmune hepatitis, dystrophy of dental enamel and nails, alopecia, vitiligo, and keratopathy and may be accompanied by hypofunction of the gonads, thyroid, pancreatic B cell, and gastric parietal cells. Autoantibodies against the cholesterol side chain cleavage enzyme (P450scc, CYP11A1) and others have been described in patients with this disorder.

The more common presentation of autoimmune adrenocortical insufficiency is associated with HLA-related disorders including type 1 diabetes mellitus and autoimmune thyroid disease. Other less common related disorders include alopecia areata, vitiligo, primary hypogonadism, pernicious anemia, and celiac disease. This disorder is often referred to as APS type 2 (APS-2). The genetic susceptibility to this disorder is linked to HLA-DR3 or DR4 (or both). These patients have anti-adrenal cytoplasmic antibodies that may be important in the pathogenesis of this disorder, and autoantibodies directed against 21α-hydroxylase (P450c21, CYP21A2) have been identified (see Chapter 2).

B. Adrenal hemorrhage Bilateral adrenal hemorrhage is a rare cause of adrenal insufficiency. The diagnosis is generally made in ill patients with significant symptoms of adrenal dysfunction. However, in some cases the diagnosis is made only after a computed tomography (CT) scan of the abdomen reveals bilateral adrenal enlargement and leads to an assessment of adrenocortical function. Anatomic factors predispose the adrenal glands to hemorrhage. The adrenal glands have a rich arterial blood supply, but they are drained by a single vein. Adrenal vein thrombosis may occur during periods of stasis or turbulence, thereby increasing adrenal vein pressure and resulting in a *vascular dam*. This causes hemorrhage into the gland and is followed by adrenocortical insufficiency.

Most patients with adrenal hemorrhage have been taking anticoagulant therapy for an underlying coagulopathy or are predisposed to thrombosis. Heparin-induced thrombocytopenia syndrome may be accompanied by adrenal vein thrombosis and
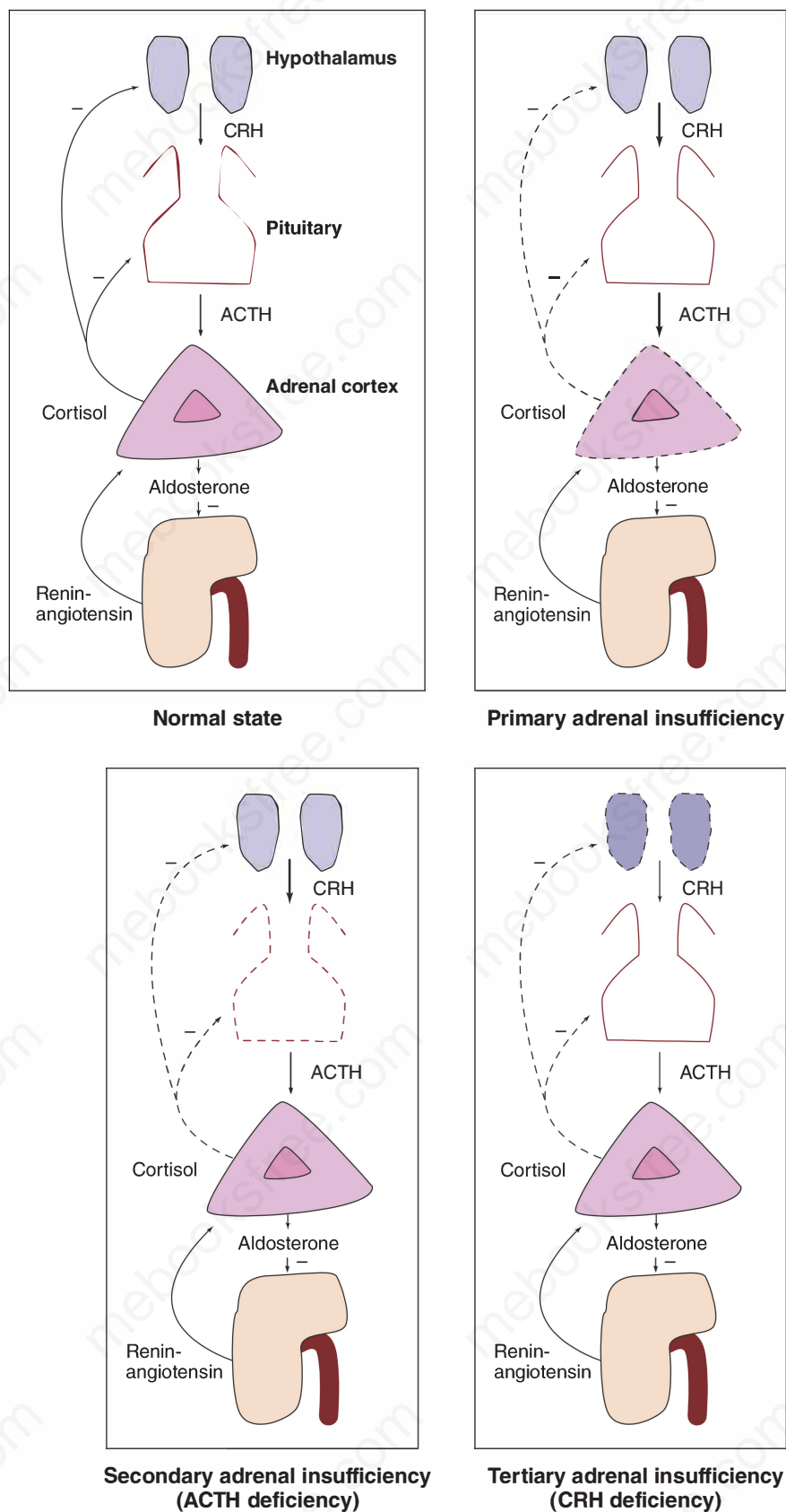
FIGURE 9–8 Hypothalamic-pituitary axis in adrenal insufficiency of different causes. These panels illustrate hormone secretion in the normal state (upper left), primary adrenal insufficiency (upper right), secondary adrenal insufficiency—ACTH deficiency (lower left), and tertiary adrenal insufficiency—CRH deficiency (lower right). Renin-angiotensin system is also illustrated. In contrast to normal secretion and hormone levels, decreased hormonal secretion is indicated by a dotted line and increased secretion by a dark solid line.

TABLE 9–3 Causes of primary adrenocortical insufficiency.

Autoimmune
Metastatic malignancy or lymphoma
Adrenal hemorrhage
Coagulopathy
Antiphospholipid antibody syndrome
Infectious
    Tuberculosis, CMV, fungi (histoplasmosis, coccidioidomycosis), HIV
Adrenoleukodystrophy
Infiltrative disorders
    Amyloidosis, hemochromatosis
Congenital adrenal hyperplasia
Familial glucocorticoid deficiency and hypoplasia
Drugs
    Ketoconazole, metyrapone, aminoglutethimide, trilostane,
        mitotane, etomidate

hemorrhage. The primary antiphospholipid antibody syndrome (lupus anticoagulant) has emerged as one of the more common causes of adrenal hemorrhage.

C. Infections   Although tuberculosis is a common cause of primary adrenal insufficiency in the rest of the world, it is a rare cause of this problem in the United States. Clinically significant adrenal insufficiency appears to occur in only about 5% of patients with disseminated tuberculosis. With the use of antituberculous chemotherapy, it may even be reversible if detected in early stages. It is important to recognize that rifampin may accelerate the metabolic clearance of cortisol, thereby increasing the replacement dose needed in these patients. Most, if not all, systemic fungal infections can involve and destroy the adrenal cortex. Tuberculosis and fungal infections are associated with enlarged adrenals that may show calcifications. Of note, the azole antifungal agents, such as ketoconazole, inhibit adrenal cytochrome P450 steroidogenic enzymes that are essential for cortisol biosynthesis. Thus, azole antifungal treatment, especially with ketoconazole, in patients with marginal adrenocortical reserve due to fungal disease may precipitate adrenal crisis.

HIV/AIDS has been associated with pathologic involvement of the adrenal gland. Although adrenal necrosis is commonly seen in postmortem examination of patients with AIDS, primary adrenal insufficiency appears to develop in few patients with this disorder. Primary adrenal insufficiency in AIDS is usually caused by opportunistic infections such as fungal infection, cytomegalovirus, and mycobacterium avium complex. Adrenocortical insufficiency usually occurs as a late manifestation in AIDS patients with very low CD4 counts (see Chapter 25).

D. Adrenoleukodystrophy   X-linked adrenoleukodystrophy is an important cause of adrenal insufficiency in men. This disorder represents two distinct entities that may cause malfunction of the adrenal cortex and demyelination in the central nervous system. These disorders are characterized by abnormally high levels of very long chain fatty acids (VLCFAs) due to their defective beta oxidation within peroxisomes. The abnormal accumulation of VLCFAs in the brain, adrenal cortex, testes, and liver result in the clinical manifestations of this disorder.

Adrenoleukodystrophy has an incidence of approximately one in 25,000 and is an X-linked disorder (chromosome Xq28) with incomplete penetrance. Molecular analysis is available clinically and can be used both in family screening and in prenatal evaluation. Two clinical phenotypes have been described. Cerebral adrenoleukodystrophy usually presents in childhood, and its neurologic symptoms include cognitive dysfunction, behavioral problems, emotional lability, and visual and gait disturbances. It may progress to dementia. Because 30% of these patients develop adrenal insufficiency before the onset of neurologic symptoms, a young man with primary adrenal insufficiency should always be screened for adrenoleukodystrophy. A clinically milder phenotype, adrenomyeloneuropathy, usually presents in the second to fourth decades of life. Spinal cord and peripheral nerve demyelination occur over years and may result in loss of ambulation, cognitive dysfunction, urinary retention, and impotence. Once again, adrenal insufficiency may occur before the onset of neurologic symptoms.

The diagnosis of adrenoleukodystrophy can be confirmed by demonstration of the defect in fatty acid metabolism with the abnormal accumulation of saturated VLCFAs, especially C26:0 fatty acid.

E. Metastatic adrenal disease   The adrenal glands are common sites of metastasis for lung, gastrointestinal, melanoma, breast, and renal neoplasia. Bilateral adrenal involvement is present in approximately 50% of patients; however, adrenal insufficiency does not occur with unilateral metastatic disease. Even with bilateral involvement adrenal insufficiency appears to be uncommon. Despite being uncommon, patients with bilateral disease should be evaluated for adrenal insufficiency and then followed as adrenal insufficiency is a potentially life threating condition. In addition, non-Hodgkin and Hodgkin lymphoma may present with involvement of the adrenal glands with bilateral adrenal enlargement and primary adrenal insufficiency.

F. Familial glucocorticoid deficiency and congenital adrenal hypoplasia   Familial glucocorticoid deficiency is a very rare disorder in which there is hereditary adrenocortical unresponsiveness to ACTH. This leads to adrenal insufficiency with subnormal glucocorticoid and adrenal androgen secretion as well as elevated plasma ACTH levels. As a rule, aldosterone secretion is preserved. At least two distinct types of this disorder have been described. One type is associated with mutations in the ACTH receptor on the cells of the adrenal cortex. Another type is often associated with achalasia and alacrima (Allgrove syndrome; triple A syndrome) and progressive neurologic impairment, but no mutations in the ACTH receptor have been seen in these patients. Mutations in the AAAS gene on chromosome 12 (12q13) are responsible for this disorder as they encode a protein called ALADIN, whose function remains incompletely understood. Congenital adrenal hypoplasia is a developmental abnormality usually presenting in the neonatal period. X-linked

forms associated with mutations of DAX1 and autosomal recessive forms have been reported.

G. Cortisol resistance   Primary cortisol resistance is an unusual disorder representing target cell resistance to cortisol due to either qualitative or quantitative abnormalities of the glucocorticoid receptor. This disorder is characterized by hypercortisolism without clinical manifestations of glucocorticoid excess. Pituitary resistance to cortisol results in hypersecretion of ACTH, which stimulates the adrenal gland to produce excessive amounts of cortisol, mineralocorticoids, and adrenal androgens. The increased production of cortisol and these nonglucocorticoid adrenal steroids may cause hypertension, hypokalemia, virilization, and sexual precocity. Because cortisol is essential for life, this disorder actually represents partial rather than complete resistance.

H. Drug-induced adrenal insufficiency   Drugs associated with primary adrenal insufficiency include the azole antifungal agents, the anesthetic agent etomidate, the antiparasitic agent suramin as well as the steroid synthesis inhibitors aminoglutethimide, metyrapone, and mitotane. Mifepristone is a progesterone antagonist that also antagonizes the glucocorticoid receptor and may cause adrenal insufficiency if given in sufficient doses. Megestrol acetate, a synthetic progesterone derivative used to stimulate appetite, binds to the glucocorticoid receptor and suppresses the HPA axis leading to adrenal insufficiency after withdrawal of therapy. Opioid narcotics cause a transient suppression of the HPA axis and can lead to suppression of plasma cortisol levels. Additionally, the use of protease inhibitors with exogenous glucocorticoids (from oral, injectable, inhaled, or even optic routes) can cause secondary adrenal insufficiency after cessation of glucocorticoid use. The immune checkpoint inhibitors are also well described to cause an array of autoimmune conditions including ACTH deficiency and secondary adrenal insufficiency.

I. Cortisol in critical illness   In 2002 it was reported that 77% of 299 patients with severe sepsis had plasma cortisol increments of less than 9 μg/dL after standard ACTH stimulation testing. These patients were classified as having *relative adrenal insufficiency* (also termed *partial adrenal insufficiency* or *decreased adrenal reserve*) and randomized to therapy with hydrocortisone and fludrocortisones with a modest survival benefit. The study led to a widespread recommendation that all severely ill patients with sepsis be tested with ACTH and treated with corticosteroids. However, the study was compromised by two major issues. First, 30% of the patients classified as relatively adrenal insufficient had received etomidate (a drug which causes adrenal insufficiency) within 12 hours of enrollment. Second, the authors did not recognize that total plasma cortisol levels and their response to ACTH are lowered in severely ill patients who frequently have hypoalbuminemia and low corticosteroid-binding globulin levels. Subsequent studies have shown that both plasma free cortisol and salivary cortisol are appropriately elevated and respond normally to ACTH in patients with sepsis. In addition, several studies have shown no survival benefit following corticosteroid therapy. Thus, there is

no convincing evidence that adrenal insufficiency is frequent in critically ill patients.

Pathophysiology

Loss of more than 90% of both adrenal cortices results in the clinical manifestations of adrenocortical insufficiency. Gradual destruction, such as occurs in the idiopathic and invasive forms of the disease, leads to chronic adrenocortical insufficiency. However, more rapid destruction occurs in many cases; about 25% of patients are in crisis or impending crisis at the time of diagnosis. With gradual adrenocortical destruction, the initial phase is that of decreased adrenal reserve; that is, basal steroid secretion is normal, but secretion does not increase in response to stress. Thus, acute adrenal crisis can be precipitated by the stresses of surgery, trauma, or infection, which require increased corticosteroid secretion. With further loss of cortical tissue, even basal secretion of glucocorticoids becomes deficient, leading to the manifestations of chronic adrenocortical insufficiency. Mineralocorticoid deficiency may occur early or late in the course. Destruction of the adrenals by hemorrhage results in sudden loss of both glucocorticoid and mineralocorticoid secretion, accompanied by acute adrenal crisis.

With decreasing cortisol secretion, plasma levels of ACTH are increased because of decreased negative feedback inhibition of their secretion. In fact, an elevation of plasma ACTH may be the earliest and most sensitive indication of suboptimal adrenocortical reserve.

Clinical Features

A. Symptoms and signs   Cortisol deficiency causes weakness, fatigue, anorexia, nausea and vomiting, hypotension, hyponatremia, and hypoglycemia. Mineralocorticoid deficiency produces renal sodium wasting and potassium retention and can lead to severe dehydration, hypotension, hyponatremia, hyperkalemia, and acidosis.

1. **Chronic primary adrenocortical insufficiency**—The chief symptoms (Table 9–4) are hyperpigmentation, weakness and fatigue, weight loss, anorexia, and gastrointestinal disturbances.
   Hyperpigmentation is the classic physical finding, and its presence in association with the above manifestations should suggest primary adrenocortical insufficiency. Generalized hyperpigmentation of the skin and mucous membranes is one of the earliest manifestations of Addison disease. It is increased

TABLE 9–4   Clinical features of primary adrenocortical insufficiency.

	Frequency
Weakness, fatigue, anorexia, weight loss	Universal
Hyperpigmentation	Universal
Hypotension	Very common
Gastrointestinal disturbances	Very common
Salt craving	Frequent
Postural symptoms	Infrequent
	Infrequent

in sun-exposed areas and accentuated over pressure areas such as the knuckles, toes, elbows, and knees. It is accompanied by increased numbers of black or dark-brown freckles. The classic hyperpigmentation of the buccal mucosa and gums is preceded by generalized hyperpigmentation of the skin; adrenal insufficiency should also be suspected when there is increased pigmentation of the palmar creases, nail beds, nipples, areolae, and perivaginal and perianal mucosa. Scars that have formed after the onset of ACTH excess become hyperpigmented, whereas older ones do not.

General weakness, fatigue and malaise, anorexia, and weight loss are invariable features of the disorder. Weight loss may reach 15 kg with progressive adrenal failure. Gastrointestinal disturbances, especially nausea and vomiting, occur in most patients; diarrhea is less frequent. An increase in gastrointestinal symptoms during an acute adrenal crisis may confuse the diagnosis by suggesting a primary intra-abdominal process.

Hypotension is present in about 90% of patients and is accompanied by orthostatic symptoms and occasionally syncope. In more severe chronic cases and in acute crises, recumbent hypotension or shock is almost invariably present. Vitiligo occurs in 4% to 17% of patients with autoimmune Addison disease but is rare in Addison disease due to other causes. Salt craving occurs in about 20% of patients.

Severe hypoglycemia may occur in children. This finding is unusual in adults but may be provoked by fasting, fever, infection, or nausea and vomiting, especially in acute adrenal crisis. Hypoglycemia occurs more commonly in secondary adrenal insufficiency.

Amenorrhea is common in Addison disease. It may be due to weight loss and chronic illness or to primary ovarian failure. Loss of axillary and pubic hair may occur in women as a result of decreased secretion of adrenal androgens.

Partial adrenal insufficiency may have subtle clinical manifestations. Adrenal insufficiency should be considered in the differential diagnosis of unexplained weight loss and in patients with hypotension and fever.

2. **Acute adrenal crisis**—Acute adrenal crisis represents a state of acute adrenocortical insufficiency and occurs in patients with Addison disease who are exposed to the stress of infection, trauma, surgery, or dehydration due to salt deprivation, vomiting, or diarrhea.

The symptoms are listed in Table 9–5. Anorexia and nausea and vomiting increase and worsen the volume depletion and dehydration. Hypovolemic shock frequently occurs, and adrenal insufficiency should be considered in any patient with unexplained vascular collapse. Abdominal pain may occur and mimic an acute abdominal emergency. Weakness, apathy, and confusion are common. Fever is also common and may be due to infection or to hypoadrenalism per se. Hyperpigmentation is present unless the onset of adrenal insufficiency is rapid and should suggest the diagnosis.

TABLE 9–5 Clinical features of acute adrenal crisis.

Hypotension and shock
Fever
Dehydration, volume depletion
Nausea, vomiting, anorexia
Weakness
Apathy, depressed mentation
Hypoglycemia

TABLE 9–6 Clinical features of adrenal hemorrhage.

	Percent
General Features	
Hypotension and shock	74
Fever	59
Nausea and vomiting	46
Confusion, disorientation	41
Tachycardia	28
Cyanosis or lividity	28
Local Features	
Abdominal, flank, or back pain	77
Abdominal or flank tenderness	38
Abdominal distention	28
Abdominal rigidity	20
Chest pain	13
Rebound tenderness	5

Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism*. 3rd ed. New York: McGraw-Hill Education; 1995.

Additional findings that suggest the diagnosis are hyponatremia, hyperkalemia, lymphocytosis, eosinophilia, and hypoglycemia.

Shock and coma may rapidly lead to death in untreated patients (see Chapter 24).

3. **Acute adrenal hemorrhage (Table 9–6)**—Bilateral adrenal hemorrhage and acute adrenal destruction in an already compromised patient with major medical illness follow a progressively deteriorating course. The usual manifestations are abdominal, flank, or back pain and abdominal tenderness. Abdominal distention, rigidity, and rebound tenderness are less frequent. Hypotension, shock, fever, nausea and vomiting, confusion, and disorientation are common; tachycardia and cyanosis are less frequent.

With progression, severe hypotension, volume depletion, dehydration, hyperpyrexia, cyanosis, coma, and death ensue.

The diagnosis of acute adrenal hemorrhage should be considered in the deteriorating patient with unexplained abdominal or flank pain, vascular collapse, hyperpyrexia, or hypoglycemia, particularly in the setting of an underlying coagulopathy.

B. Laboratory and electrocardiographic findings and imaging studies

1. **Gradual adrenal destruction**—Hyponatremia and hyperkalemia are classic manifestations of the glucocorticoid and mineralocorticoid deficiency of primary adrenal insufficiency and should suggest the diagnosis. Hematologic manifestations include normocytic, normochromic anemia, neutropenia, eosinophilia, and a relative lymphocytosis. Azotemia with increased concentrations of blood urea nitrogen and serum creatinine is due to volume depletion and dehydration. Mild acidosis is frequently present. Hypercalcemia of mild to moderate degree occurs in about 6% of patients.

Abdominal radiographs reveal adrenal calcification in half the patients with tuberculous Addison disease and in some patients with other invasive or hemorrhagic causes of adrenal insufficiency. CT is a more sensitive detector of adrenal calcification and adrenal enlargement. Bilateral adrenal enlargement in association with adrenal insufficiency may be seen with tuberculosis, fungal infections, cytomegalovirus, malignant and nonmalignant infiltrative diseases, and adrenal hemorrhage.

Electrocardiographic features show low voltage, a vertical QRS axis, and nonspecific ST-T wave abnormalities secondary to abnormal electrolytes.

2. **Acute adrenal hemorrhage**—Hyponatremia and hyperkalemia occur in only a small number of cases, but azotemia is a common finding. Increased circulating eosinophils may suggest the diagnosis. The diagnosis is frequently established only when imaging studies reveal bilateral adrenal enlargement.

SECONDARY ADRENOCORTICAL INSUFFICIENCY

Etiology

Secondary adrenocortical insufficiency due to ACTH deficiency is most commonly a result of exogenous glucocorticoid therapy. Pituitary or hypothalamic tumors, and their treatment, are the most common causes of naturally occurring pituitary ACTH hyposecretion. These and other less common causes are reviewed in Chapter 4.

Pathophysiology

ACTH deficiency is the primary event and leads to decreased cortisol and adrenal androgen secretion. Aldosterone secretion almost always remains normal as renin continues to stimulate aldosterone production. In the early stages, basal ACTH and cortisol levels may be normal; however, ACTH reserve is impaired, and ACTH and cortisol responses to stress are, therefore, subnormal. With further loss of basal ACTH secretion, there is atrophy of the zonae fasciculata and reticularis of the adrenal cortex; and, therefore, basal cortisol secretion is decreased. At this stage, the entire pituitary adrenal axis is impaired; that is, there is not only decreased ACTH responsiveness to stress but also decreased adrenal responsiveness to acute stimulation with exogenous ACTH.

The manifestations of glucocorticoid deficiency are similar to those described for primary adrenocortical insufficiency. However, because aldosterone secretion by the zona glomerulosa is usually preserved, the manifestations of mineralocorticoid deficiency are absent.

Clinical Features

A. Symptoms and signs   Secondary adrenal insufficiency is usually chronic, and the manifestations may be nonspecific. However, acute crisis can occur in undiagnosed patients or in corticosteroid-treated patients who do not receive increased steroid dosage during periods of stress.

The clinical features of secondary adrenal insufficiency differ from those of primary adrenocortical insufficiency, in that pituitary secretion of ACTH is deficient, and hyperpigmentation is, therefore, not present. In addition, mineralocorticoid secretion is usually normal. Thus, the clinical features of ACTH and glucocorticoid deficiency are nonspecific.

Volume depletion, dehydration, and hyperkalemia are usually absent. Hypotension is usually not present except in acute presentations. Hyponatremia may occur, as a result of water retention

and inability to excrete a water load, but is not accompanied by hyperkalemia. Prominent features are weakness, lethargy, easy fatigability, anorexia, nausea, and occasionally vomiting. Arthralgias and myalgias also occur. Hypoglycemia is occasionally the presenting feature. Acute decompensation with severe hypotension or shock unresponsive to vasopressors may occur.

B. Associated features   Patients with secondary adrenal insufficiency commonly have additional features that suggest the diagnosis. A history of glucocorticoid therapy, or if this is not available, the presence of Cushingoid features suggests prior glucocorticoid use. Hypothalamic or pituitary tumors leading to ACTH deficiency usually cause loss of other pituitary hormones (hypogonadism and hypothyroidism). Hypersecretion of GH or prolactin (PRL) from a pituitary adenoma may be present.

C. Laboratory findings   Findings on routine laboratory examination consist of normochromic, normocytic anemia, neutropenia, lymphocytosis, and eosinophilia. Hyponatremia is not uncommon and may be the presenting laboratory abnormality. Hyponatremia is due to the lack of glucocorticoid-negative feedback on AVP as well as the reduction in glomerular filtration associated with hypocortisolism. Serum potassium, creatinine, and bicarbonate and blood urea nitrogen are usually normal; plasma glucose may be low, although severe hypoglycemia is unusual.

DIAGNOSIS OF ADRENOCORTICAL INSUFFICIENCY

Although the diagnosis of adrenal insufficiency should be confirmed by assessment of the pituitary adrenal axis, therapy should not be delayed, and the patient should not be subjected to procedures that may increase volume loss and dehydration and further contribute to hypotension. If the patient is acutely ill, therapy should be instituted and the diagnosis established when the patient is stable.

Diagnostic Tests

Because basal levels of adrenocortical steroids in either urine or plasma may be normal in partial adrenal insufficiency, tests of adrenocortical reserve are necessary to establish the diagnosis (Figure 9–9). These tests are described in the section on laboratory evaluation and in Chapter 4.

Rapid ACTH Stimulation Test

The rapid ACTH stimulation test assesses adrenal reserve and is the initial procedure in the assessment of possible adrenal insufficiency, either primary or secondary. As previously discussed, the low-dose ACTH (1-μg cosyntropin) stimulation test has been suggested to represent a more physiologic stimulus to the adrenal cortex; however, evidence-based guidelines for the diagnosis of adrenal insufficiency suggest that the standard 250 μg cosyntropin dose provides the best diagnostic certainty, and the low-dose
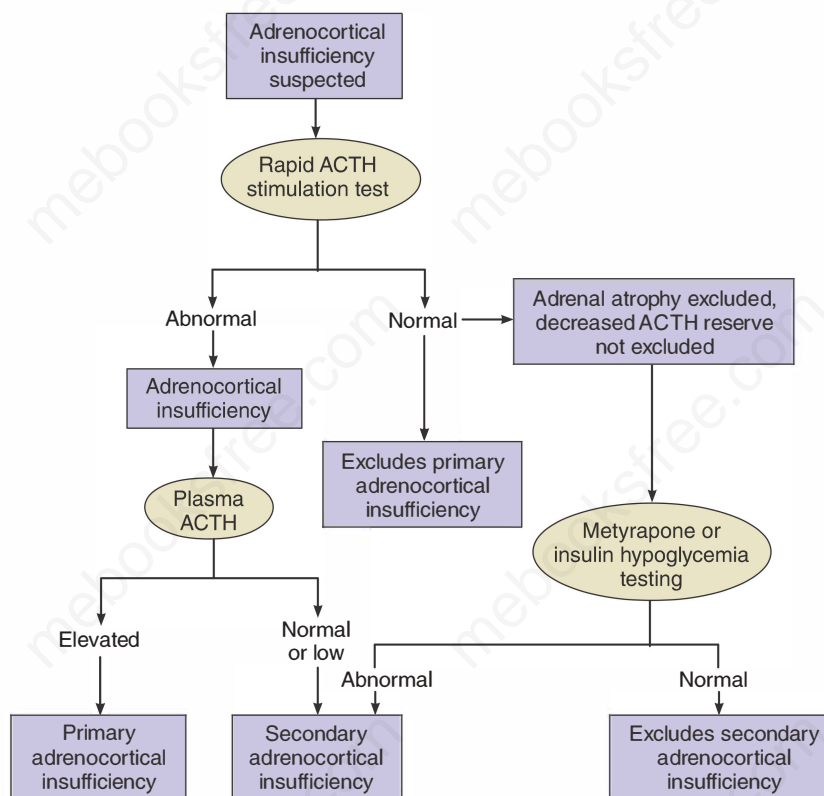
FIGURE 9–9 Evaluation of suspected primary or secondary adrenocortical insufficiency. Boxes enclose clinical decisions, and circles enclose diagnostic tests. (Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism*. 3rd ed. New York: McGraw-Hill Education; 1995.)

(1 μg) test should be reserved for centers with well-validated normal controls and extensive clinical experience.

Subnormal responses to exogenous ACTH administration are an indication of decreased adrenal reserve and establish the diagnosis of adrenocortical insufficiency. Further diagnostic procedures are not required, because subnormal responses to the rapid ACTH stimulation test indicate lack of responsiveness to metyrapone, insulin-induced hypoglycemia, or stress. However, this test does not permit differentiation of primary and secondary causes. This is best accomplished by measurement of basal plasma ACTH levels, as discussed later.

A normal response to the rapid ACTH stimulation test excludes primary adrenal failure, because a normal cortisol response indicates normal cortical function. However, normal responsiveness does not exclude partial secondary adrenocortical insufficiency in those few patients with decreased pituitary reserve and decreased stress responsiveness of the HPA axis who maintain sufficient basal ACTH secretion to prevent adrenocortical atrophy. If this situation is suspected clinically, pituitary ACTH responsiveness may be tested directly with metyrapone, CRH stimulation, or insulin-induced hypoglycemia (see section "Laboratory Evaluation" and Table 9–7).

TABLE 9–7 ACTH and cortisol response in adrenal insufficiency of varying etiologies.

	Basal State		Response to CRH Infusion	
	ACTH	**Cortisol**	**ACTH**	**Cortisol**
Normal Adrenal Function	Normal	Normal	Significant Increase	Significant Increase
Acute Primary AI	Very High	Low	Increase	No Change
Chronic Primary AI	Very High	Low	Increase	No Change
Acute Secondary AI	Normal to Low	Low	No Change or Small Increase	No Change or Small Increase
Chronic Secondary AI	Low	Low	No Change	No Change

Plasma ACTH Levels

If adrenal insufficiency is present, plasma ACTH levels are used to differentiate primary and secondary forms. In patients with primary adrenal insufficiency, plasma ACTH levels exceed the upper limit of the reference range (>52 pg/mL [11 pmol/L]) and usually exceed 200 pg/mL (44 pmol/L). Plasma ACTH concentrations are inappropriately normal or less than 10 pg/mL (2.2 pmol/L) in patients with secondary adrenal insufficiency (see Figure 9–9). However, basal ACTH levels must always be interpreted in light of the clinical situation, especially because of the episodic nature of ACTH secretion and its short plasma half-life. For example, ACTH levels frequently exceed the normal range during the recovery of the HPA axis from secondary adrenal insufficiency and may be confused with levels seen in primary adrenal insufficiency. Patients with primary adrenal insufficiency consistently have elevated ACTH levels. In fact, the ACTH concentration is elevated early in the course of adrenal insufficiency before a significant reduction in the basal cortisol level or its response to exogenous ACTH occurs. Therefore, plasma ACTH measurements serve as a valuable screening study for primary adrenal insufficiency.

Partial ACTH Deficiency

When partial ACTH deficiency and decreased pituitary reserve are suspected despite normal responsiveness to the rapid ACTH stimulation test, the following procedures may be used for more direct assessment of hypothalamic-pituitary function.

A. Methods of testing   The overnight metyrapone test is used in patients with suspected hypothalamic or pituitary disorders when hypoglycemia is contraindicated and in those with prior glucocorticoid therapy. Insulin-induced hypoglycemia is used in patients with suspected hypothalamic or pituitary tumors, because both ACTH and GH responsiveness can be assessed (see Chapter 4).

B. Interpretation   A normal response to either metyrapone or hypoglycemia excludes secondary adrenocortical insufficiency (see section "Laboratory Evaluation"). Subnormal responses, even in the presence of a normal response to ACTH administration, establish the diagnosis of secondary adrenal insufficiency.

TREATMENT OF ADRENOCORTICAL INSUFFICIENCY

The aim of treatment of adrenocortical insufficiency is to produce levels of glucocorticoids and mineralocorticoids equivalent to those achieved in an individual with normal HPA function under similar circumstances.

Acute Addisonian Crisis (Table 9–8)

Treatment for acute Addisonian crisis should be instituted as soon as the diagnosis is suspected and not delayed until the diagnosis is

TABLE 9–8   Treatment of acute adrenal crisis.

Glucocorticoid Replacement
(1) Administer hydrocortisone sodium phosphate or sodium succinate, 50-100 mg intravenously every 6-12 h for 24 h
(2) When the patient is stable, reduce the dosage by 50%
(3) Taper to maintenance therapy on day 3-5 and add mineralocorticoid therapy as required

General and Supportive Measures
(1) Correct volume depletion, dehydration, and hypoglycemia with intravenous saline and glucose
(2) Evaluate and correct infection and other precipitating factors

confirmed. Therapy includes administration of glucocorticoids; correction of dehydration, hypovolemia, and electrolyte abnormalities; general supportive measures; and treatment of coexisting or precipitating disorders.

A. Cortisol (Hydrocortisone)   Parenteral cortisol in soluble form (hydrocortisone hemisuccinate or phosphate) is the glucocorticoid preparation most commonly used. When administered in supraphysiologic doses, hydrocortisone has sufficient sodium-retaining potency so that additional mineralocorticoid therapy is not required in patients with primary adrenocortical insufficiency.

Cortisol in intravenous doses of 50 to 100 mg is given every 6 to 12 hours for the first 24 hours. The response to therapy is usually rapid, with improvement occurring within 12 hours or less. If improvement occurs and the patient is stable, doses can be weaned by 50% on the second day, and in most patients the dosage may then be gradually reduced to maintenance dose thereafter (see section "Maintenance Therapy," later).

1. In severely ill patients, especially in those with additional major complications (eg, sepsis), higher cortisol doses (100 mg intravenously every 6-8 hours) are maintained until the patient is stable.
2. In primary Addison disease, mineralocorticoid replacement, in the form of fludrocortisone (see later), is usually added when hydrocortisone replacement is reduced to near physiological levels.
3. In secondary adrenocortical insufficiency with acute crisis, the primary requirement is glucocorticoid replacement and is satisfactorily supplied by the administration of cortisol, as outlined earlier. If the possibility of excessive fluid and sodium retention in such patients is of concern, equivalent parenteral doses of synthetic steroids such as prednisolone or dexamethasone may be substituted.

B. Intravenous fluids   Intravenous glucose and saline are administered to correct volume depletion, hypotension, and hypoglycemia. Volume deficits may be severe in Addison disease, and hypotension and shock may not respond to vasopressors unless glucocorticoids are administered. Hyperkalemia and acidosis are usually corrected with cortisol and volume replacement; however, an occasional patient may require specific therapy for these abnormalities (see also Chapter 24).

TABLE 9–9 **Regimen for maintenance therapy of primary adrenocortical insufficiency.**

(1) Hydrocortisone, 10-15 mg in AM, 5-10 mg in the afternoon (4-5 PM)

(2) Fludrocortisone, 0.05-0.1 mg orally in AM

(3) Clinical follow-up: maintenance of normal weight, blood pressure, and electrolytes with regression of clinical features

(4) Patient education plus identification card, medical alert bracelet or necklace (*dog tag*). Useful patient education materials are available from the National Institutes of Health. http://www.cc.nih.gov/ccc/patient_education/pepubs/mngadrins.pdf

(5) Increase hydrocortisone dosage during stress. Double or triple the oral dose for mild illness. Provide patient with injectable form of glucocorticoid for emergency use

Maintenance Therapy (Table 9–9)

Patients with Addison disease require life-long glucocorticoid and mineralocorticoid therapy. Cortisol (hydrocortisone) is the glucocorticoid preparation of choice. The basal production rate of cortisol is approximately 8 to 12 mg/m^2/d. The maintenance dose of hydrocortisone is usually 15 to 25 mg daily in adults. The oral dose is usually divided into 10 to 15 mg in the morning on arising and 5 to 10 mg later in the day. Cortisol in twice-daily doses gives satisfactory responses in most patients; however, some patients may require only a single morning dose, and others may require three doses daily (eg, 10-5-5) to maintain well-being and normal energy levels. Insomnia is a side-effect of glucocorticoid administration and can usually be prevented by administering the last dose in the early afternoon.

Fludrocortisone (9α-fluorocortisol) is used for mineralocorticoid therapy; the usual doses are 0.05 to 0.2 mg/d orally in the morning. Because of the long half-life of this agent, divided doses are not required. About 10% of addisonian patients can be managed with cortisol and adequate dietary sodium intake alone and do not require fludrocortisone.

Secondary adrenocortical insufficiency is treated with the cortisol dosages described earlier for the primary form. Fludrocortisone is rarely required. The recovery of normal function of the HPA axis following suppression by exogenous glucocorticoids may take weeks to years, and its duration is not readily predictable. Consequently, prolonged replacement therapy may be required. Several studies in women have suggested potential benefits of DHEA in doses of 50 mg/d in terms of improvement in well-being, although this remains controversial.

Response to Therapy

General clinical signs, such as good appetite and sense of well-being, are the guides to the adequacy of replacement. Obviously, signs of Cushing syndrome indicate overtreatment. It is generally expected that the daily dose of hydrocortisone should be doubled or tripled during periods of minor stress, and the dose needs to be increased to as much as 200 to 300 mg/d during periods of major stress, such as a surgical procedure. Acute adrenal crisis is a common problem in patients with established adrenal insufficiency (8%-10% per year); accordingly, appropriate sick day

management with stress dosing of glucocorticoids should be reinforced frequently. Patients receiving excessive doses of glucocorticoids are also at risk for increased bone loss and clinically significant osteoporosis. Therefore, the replacement dose of glucocorticoid should be maintained at the lowest amount needed to provide the patient with a proper sense of well-being. Two major factors have prompted a reassessment of the issue of appropriateness and possible excess glucocorticoid replacement for patients with adrenal insufficiency. First, there is a greater appreciation of the potential risks of overtreatment or undertreatment. Recent evidence suggests that subclinical Cushing syndrome associated with adrenal incidentaloma contributes to elevated blood sugar and blood pressure, decreased bone mineral density, increased cardiovascular disease, and even increased mortality. The levels of cortisol secretion by many incidentalomas are similar to those observed in patients with adrenal insufficiency receiving mild cortisol overreplacement. In addition, studies in patients receiving glucocorticoid replacement therapy demonstrate an inverse relationship between dose and bone mineral density and a positive correlation between dose and markers of bone resorption. Second, there is recognition that there is considerable variation among individuals in terms of the plasma levels of cortisol achieved with orally administered hydrocortisone or cortisol.

The measurement of urine free cortisol does not provide a reliable index for appropriate glucocorticoid replacement. Similarly, ACTH measurements are not a good indication of the adequacy of glucocorticoid replacement; marked elevations of plasma ACTH in patients with chronic adrenal insufficiency are often not suppressed into the normal range despite adequate hydrocortisone replacement. In fact, plasma ACTH values that fall to low normal or low levels suggest overreplacement with glucocorticoids. Plasma cortisol day curves—multiple samples for plasma cortisol concentration—have been proposed but not yet widely adopted. Therefore, it is prudent to use clinical assessment of the patient's response to therapy, rather than biochemical criteria, to gauge the adequacy of steroid replacement.

Adequate treatment results in the disappearance of weakness, malaise, and fatigue. Anorexia and other gastrointestinal symptoms resolve, and weight returns to normal. The hyperpigmentation invariably improves but may not entirely disappear. Inadequate cortisol administration leads to persistence of these symptoms of adrenal insufficiency, and excessive pigmentation remains.

Adequate mineralocorticoid replacement may be determined by assessment of blood pressure and electrolyte composition. With adequate treatment, the blood pressure is normal without orthostatic change, and serum sodium and potassium remain within the normal range. Some endocrinologists monitor plasma renin activity (PRA) as an objective measure of fludrocortisone replacement. Upright PRA levels are usually less than 5 ng/mL/h in adequately replaced patients. Hypertension and hypokalemia result if the fludrocortisone dose is excessive. Conversely, undertreatment may lead to fatigue and malaise, orthostatic symptoms, and subnormal supine or upright blood pressure, with hyperkalemia and hyponatremia.

Prevention of Adrenal Crisis

The development of acute adrenal insufficiency in previously diagnosed and treated patients is almost entirely preventable in compliant individuals. The essential elements are patient education and increased glucocorticoid dosages during illness.

The patient should be informed about the necessity for lifelong therapy, the possible consequences of acute illness, and the necessity for increased therapy and medical assistance during acute illness. An identification card or bracelet should be carried or worn at all times.

The cortisol dose should be increased by the patient to 60 to 80 mg/d with the development of minor illnesses; the usual maintenance dosage may be resumed in 24 to 48 hours if improvement occurs. Increased mineralocorticoid therapy is not required.

If symptoms persist or become worse, the patient should continue increased cortisol doses and call the physician.

Vomiting may result in inability to ingest or absorb oral cortisol, and diarrhea in addisonian patients may precipitate a crisis because of rapid fluid and electrolyte losses. Patients must understand that if these symptoms occur, they should seek immediate medical assistance so that parenteral glucocorticoid therapy can be given.

Steroid Coverage for Surgery (Table 9–10)

The normal physiologic response to surgical stress involves an increase in cortisol secretion. The increased glucocorticoid activity may serve primarily to modulate the immunologic response to stress. Thus, patients with primary or secondary adrenocortical insufficiency scheduled for elective surgery require increased glucocorticoid coverage. This problem is most frequently encountered in patients with pituitary-adrenal suppression due to exogenous glucocorticoid therapy. The principles of management are outlined in Table 9–10.

PROGNOSIS OF ADRENOCORTICAL INSUFFICIENCY

Before glucocorticoid and mineralocorticoid therapy became available, primary adrenocortical insufficiency was invariably fatal, with death usually occurring within 2 years after onset. Survival

TABLE 9–10 **Steroid coverage for surgery.**

(1) Correct electrolytes, blood pressure, and hydration if necessary

(2) Give hydrocortisone sodium phosphate or sodium succinate, 100 mg intramuscularly, on call to operating room

(3) Give 50 mg intramuscularly or intravenously in the recovery room and then every 6-8 h for the first 24 h

(4) If progress is satisfactory, reduce dosage to 25 mg every 6-8 h for 24 h and then taper to maintenance dosage over 2-5 d. Resume previous fludrocortisone dose when the patient is taking oral medications

(5) Maintain or increase hydrocortisone dosage to 200-400 mg/d if fever, hypotension, or other complications occur

Reproduced with permission from Felig P, Baxter JD, Frohman LA. *Endocrinology and Metabolism.* 3rd ed. New York: McGraw-Hill Education; 1995.

now depends on the underlying cause of the adrenal insufficiency. In patients with autoimmune Addison disease, survival approaches that of the normal population, and most patients lead normal lives. In general, death from adrenal insufficiency now occurs only in patients with rapid onset of disease who may die before the diagnosis is established and appropriate therapy started.

Secondary adrenal insufficiency has an excellent prognosis with glucocorticoid therapy. Adrenal insufficiency due to bilateral adrenal hemorrhage may be fatal, with some cases being recognized only at autopsy.

CUSHING SYNDROME

Chronic glucocorticoid excess, whatever its cause, leads to the constellation of symptoms and physical features known as **Cushing syndrome**. It is most commonly iatrogenic, resulting from chronic glucocorticoid therapy. Spontaneous Cushing syndrome is caused by abnormalities of the pituitary or adrenal gland or may occur as a consequence of ACTH or CRH secretion by nonpituitary tumors (**ectopic ACTH syndrome; ectopic CRH syndrome**) (Figure 9–10). **Cushing disease** is defined as the specific type of Cushing syndrome due to excessive pituitary ACTH secretion from a pituitary tumor. This section reviews the various types of endogenous Cushing syndrome and discuss their diagnosis and therapy (see also Chapter 4).

Classification and Incidence

Cushing syndrome is classified as either ACTH dependent or ACTH independent (Table 9–11). The ACTH-dependent types of Cushing syndrome—ectopic ACTH syndrome and Cushing disease—are characterized by chronic ACTH hypersecretion, which results in hyperplasia of the adrenal zonae fasciculata and reticularis and, therefore, increased secretion of cortisol, androgens, and DOC.

ACTH-independent Cushing syndrome may be caused by a primary adrenal neoplasm (adenoma or carcinoma) or bilateral nodular adrenal hyperplasia. In these cases, the cortisol excess suppresses pituitary ACTH secretion.

A. Cushing disease   Traditionally this has been the most frequent type of Cushing syndrome and most studies attribute 80% of reported cases to Cushing disease. It is much more common in women than in men (female-male ratio of about 5:1) and the age at diagnosis is usually 20 to 40 years but may range from childhood to 70 years.

B. Ectopic ACTH hypersecretion   This disorder accounts for approximately 10% of patients with ACTH-dependent Cushing syndrome. The production of ACTH from a tumor of nonpituitary origin may result in severe hypercortisolism, but many of these patients lack the classic features of glucocorticoid excess. This presumably reflects the acuteness of the clinical course in the ectopic ACTH syndrome. The clinical
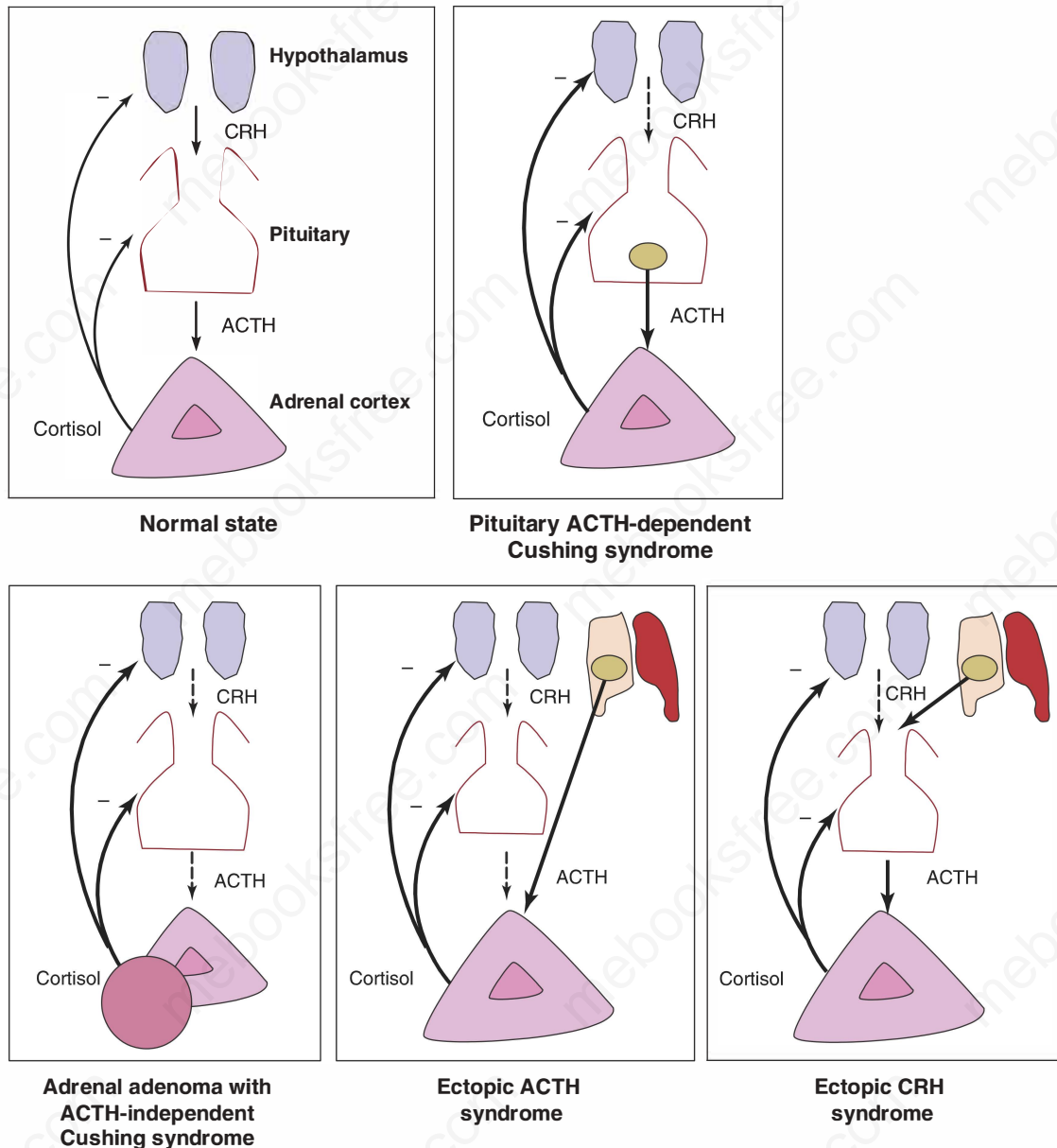
FIGURE 9–10 Hypothalamic-pituitary axis in Cushing syndrome of different causes. These panels illustrate hormone secretion in the normal state (upper left), and four types of cortisol excess: Pituitary ACTH-dependent (with an ACTH-secreting pituitary tumor) (upper right), adrenal tumor (lower left), ectopic ACTH syndrome due to an ACTH-secreting lung cancer (lower middle), and ectopic CRH syndrome due to a CRH-secreting lung tumor. In contrast to normal secretion and hormone levels, decreased hormonal secretion is indicated by a dotted line and increased secretion by a dark solid line.

presentation of ectopic ACTH secretion is most frequently seen in patients with tumors of a thoracic origin. Bronchial carcinoid, small cell and non-small cell lung carcinoma are responsible for over 50% of cases of this syndrome. The prognosis in patients with the ectopic ACTH syndrome is generally poor and largely a function of the primary tumor. Individuals with ectopic ACTH syndrome from small cell lung carcinoma have a mean survival of less than 12 months. The ectopic ACTH syndrome may also present in a fashion identical to classic Cushing disease and pose a challenging diagnostic dilemma. A wide

variety of tumors has been reported to produce ectopic ACTH and may be radiologically inapparent at the time of the presentation. The ectopic ACTH syndrome is more common in men, and the peak age incidence is 40 to 60 years.

C. Primary adrenal tumors   Primary adrenal tumors are increasingly being recognized as a common cause of Cushing syndrome. Most of these patients have benign adrenocortical adenomas. Adrenocortical carcinomas are uncommon, with an incidence of approximately 2 per million per year. Adrenal

TABLE 9–11 Cushing syndrome: differential diagnosis.

ACTH-Dependent
Pituitary adenoma (Cushing disease)
Nonpituitary neoplasm (ectopic ACTH)
ACTH-Independent
Iatrogenic (glucocorticoid, megestrol acetate)
Adrenal carcinoma
Unilateral adrenal neoplasm/adenoma
Bilateral adrenal hyperplasia/neoplasm
Bilateral macronodular adrenal hyperplasia
McCune Albright syndrome
Primary pigmented nodular adrenal disease
Isolated micronodular adrenocortical disease

carcinomas are more common in women. Autonomous cortisol secretion without the classic features of Cushing syndrome occurs in up to 20% to 25% of patients with adrenal incidentalomas (see later).

D. Childhood Cushing syndrome   Childhood Cushing disease is rare, but is more common in the adolescent population than in younger children. Most pediatric cases occur in patients over 10 years of age with an equal distribution between the sexes. Cushing disease remains the most common cause of Cushing syndrome accounting for more than 80% of cases. Primary adrenal disease accounts for an additional 10% to 15% of all pediatric cases. A hallmark of childhood Cushing syndrome is significant weight gain without accompanying linear growth.

Pathology

A. Anterior pituitary gland

1. **Pituitary adenomas**—Pituitary adenomas are present in over 90% of patients with Cushing disease. These tumors are typically smaller than those secreting GH or PRL; 80% to 90% are less than 10 mm in diameter. A small group of patients have larger tumors (>10 mm); these macroadenomas are frequently invasive, leading to extension outside the sella turcica. Malignant pituitary tumors occur very rarely.

   Microadenomas are located within the anterior pituitary; they are not encapsulated but surrounded by a rim of compressed normal anterior pituitary cells. With routine histologic stains, these tumors are composed of compact sheets of well-granulated basophilic cells in a sinusoidal arrangement. ACTH, β-lipotropin, and β-endorphin have been demonstrated in these tumor cells by immunocytochemical methods. Larger tumors may appear chromophobic on routine histologic study; however, they also contain ACTH and its related peptides. These ACTH-secreting adenomas typically show Crooke changes (a zone of perinuclear hyalinization that is the result of chronic exposure of corticotroph cells to hypercortisolism). Electron microscopy demonstrates secretory granules that vary in size from 200 to 700 nm. The number of granules varies in individual cells; they may be dispersed throughout the cytoplasm or concentrated along the cell membrane. A typical feature of these adenomas is the presence of bundles of perinuclear microfilaments (average 7 nm in diameter)

surrounding the nucleus; these are responsible for Crooke hyaline changes visible on light microscopy.

2. **Hyperplasia**—Diffuse hyperplasia of corticotroph cells has been reported rarely in patients with Cushing disease.

3. **Other conditions**—In patients with adrenal tumors or ectopic ACTH syndrome, the pituitary corticotrophs show prominent Crooke hyaline changes and perinuclear microfilaments. The ACTH content of corticotroph cells is reduced consistent with their suppression by excessive cortisol secretion present in these conditions.

B. Adrenocortical hyperplasia

Chronic ACTH hypersecretion, usually due to Cushing disease, can lead to bilateral enlargement and hyperplasia of the adrenal cortex. Combined adrenal weight (normal, 8-10 g) is modestly increased, ranging from 12 to 24 g. On histologic study, there is equal hyperplasia of the compact cells of the zona reticularis and the clear cells of the zona fasciculata; consequently, the width of the cortex is increased. Electron microscopy reveals normal ultrastructural features. When ACTH levels are very high as in the ectopic ACTH syndrome, the adrenals are frequently larger, with combined weights up to or more than 50 g. The characteristic microscopic feature is marked hyperplasia of the zona reticularis; columns of compact reticularis cells expand throughout the zona fasciculata and into the zona glomerulosa. The zona fasciculata clear cells are markedly reduced.

C. Nodular adrenal disease

Adrenal nodules that hypersecrete cortisol can be unilateral or bilateral, and their pathology varies depending on the etiology.

1. **Unilateral cortisol-secreting adrenal adenomas**—These adenomas are encapsulated, usually weigh 10 to 70 g, and range in size from 1 to 6 cm. Microscopically, clear cells of the zona fasciculata type predominate, although cells typical of the zona reticularis are also seen. The uninvolved adrenal cortex contiguous to the tumor and that of the contralateral gland are atrophic in the presence of functioning adrenal adenomas and carcinomas (see later). The cortex is markedly thinned, whereas the capsule is thickened. Histologically, the zona reticularis is virtually absent; the remaining cortex is composed of clear fasciculata cells. The architecture of the zona glomerulosa is normal.

2. **Adrenal carcinomas**—Adrenal carcinomas are usually greater than 4 cm when diagnosed and often weigh more than 100 g, occasionally exceeding 1 kg. They may be palpable as abdominal masses. Grossly, they are encapsulated and highly vascular; necrosis, hemorrhage, and cystic degeneration are common, and areas of calcification may be present. The histologic appearance of these carcinomas varies considerably; they may appear to be benign or may exhibit considerable pleomorphism. Vascular or capsular invasion is predictive of malignant behavior, as is local extension. These carcinomas invade local structures (kidney, liver, and retroperitoneum) and metastasize hematogenously to liver and lung. As in unilateral glucocorticoid-secreting adenomas, the surrounding adrenal cortex is atrophic.

3. **Bilateral nodular adrenal hyperplasia**—This can be due to ACTH stimulation (ACTH-dependent) or from several unique

ACTH-independent pathophysiologic disorders. Longstanding ACTH hypersecretion—either pituitary or nonpituitary—may result in nodular enlargement of the adrenal gland. The cortex between the nodules is not atrophic as there is ACTH stimulation to the entire gland. These focal nodules are often mistaken for adrenal neoplasms and may lead to unnecessary as well as unsuccessful unilateral adrenal surgery. In occasional cases these nodules may, over a period of time, even become autonomous or semiautonomous. Removal of the ACTH-secreting neoplasm results in regression of the adrenal nodules as well as resolution of hypercortisolism unless the nodules have already developed significant autonomy.

ACTH-independent bilateral nodular adrenal hyperplasia can be divided based on nodular size, diameter less than 1 cm (micronodular) or more than 1 cm (macronodular). Primary pigmented nodular adrenocortical disease (PPNAD) is a micronodular condition that has small to normal sized adrenal glands with multiple black and brown nodules with intranodular cortical atrophy. The pathology of nodules in bilateral macronodular adrenal hyperplasia (BMAH) is similar to unilateral disease, but the internodular cortex can be hyperplastic (commonly) or can be atrophic.

Pathogenesis and Genetics (Table 9–12)

A. Cushing disease The causes, natural history, and genetics of Cushing disease are reviewed in Chapter 4. Current evidence is consistent with the view that spontaneously arising corticotroph-cell pituitary adenomas are the primary cause and that the consequent ACTH hypersecretion and hypercortisolism lead to the characteristic endocrine abnormalities and hypothalamic dysfunction. This is supported by evidence showing that selective removal of these adenomas by pituitary microsurgery reverses the abnormalities and is followed by return of the HPA axis to normal. In addition, molecular studies have shown that nearly all corticotroph adenomas are monoclonal.

Although these primary pituitary adenomas are responsible for the great majority of cases, a few patients have been described in whom pituitary disease has been limited to corticotroph-cell

hyperplasia; these may be secondary to excessive CRH secretion by rare, benign hypothalamic gangliocytoma.

B. Ectopic ACTH syndrome and ectopic CRH syndrome
Ectopic ACTH syndrome arises when nonpituitary tumors synthesize and secrete biologically active ACTH. The related peptides β-lipotropin and β-endorphin are also synthesized and secreted, as are inactive ACTH fragments. Production of CRH has also been demonstrated in ectopic tumors secreting ACTH, but whether CRH plays a role in pathogenesis is unclear. A few cases in which nonpituitary tumors produced only CRH have been reported.

Ectopic ACTH syndrome occurs predominantly in selected tumor types (Figure 9–11); carcinoid tumors of lung, non-small cell, and small cell carcinoma of the lung account for most of the cases. Other tumors causing the syndrome are of neuroendocrine origin from cells in the thymus, gut, ovary, pancreatic islet, or thyroid (ie, C-cells). Multiple other tumor types have also been reported to cause the ectopic ACTH syndrome, but are very rare causes for it (see Chapter 21).

C. Nodular adrenal disease Recent discoveries have allowed a better understanding of the genetics responsible for nodular adrenal disease.

1. **Unilateral cortisol-secreting adrenal adenomas**—In 2013, four separate groups from around the world reported a mutation in the gene for the protein kinase A catalytic subunit, PRKACA. Together, these groups investigated more than 200 cortisol-secreting adrenal adenomas. They found mutations in the gene for PRKACA leading to constitutive activation of protein kinase A in more than 40% of the adenomas. The mutations were largely (95%) found in a single hotspot. These mutations were not found to surrounding normal adrenal tissue, in non-functional adenomas, adrenal cortical carcinomas, or in adrenal nodules with subtle cortisol hypersecretion (subclinical Cushing syndrome).

2. **Bilateral micronodular adrenal hyperplasia**—Primary pigmented nodular adrenocortical disease is an uncommon

TABLE 9–12 Genetic causes of adrenal dependent cushing syndrome.

Condition/Gene/Chromosome	Related Conditions	Nodular Disease
McCune Albright syndrome/GNAS1/20q13.3	Fibrous dysplasia of bone, café-au-lait skin pigmentation, precocious puberty	Unilateral or bilateral, micro- or macronodularity
Primary pigmented nodular adrenal disease/PRKAR1A/17q22-24	Carney Complex (myxomas, testicular tumors, thyroid nodules, growth hormone producing pituitary adenomas)	Bilateral, micronodular
Cortisol producing adenomas/PRKACA/19p13.1	No associated syndrome	Unilateral, usually macronodular
Familial adenomatous polyposis/APC/5q22.2	Multiple colon polyps and colon cancer	Unilateral or bilateral, macronodular
Multiple endocrine neoplasia type 1/MENIN/11q13	Primary hyperparathyroidism, pancreatic endocrine tumors, pituitary adenomas	Unilateral or bilateral, macronodular
Hereditary leiomyomatosis and renal cell cancer syndrome/fumarate hydratase/1q42.1	Hereditary leiomyomatosis, renal cell cancer, uterine fibroids	Bilateral, macromodular
Hereditary bilateral adrenal adenomas/ARMC5/16q11.2	Hereditary adrenal adenomas, often cortisol-producing (may be subclinical cortisol excess)	Bilateral, macronodular
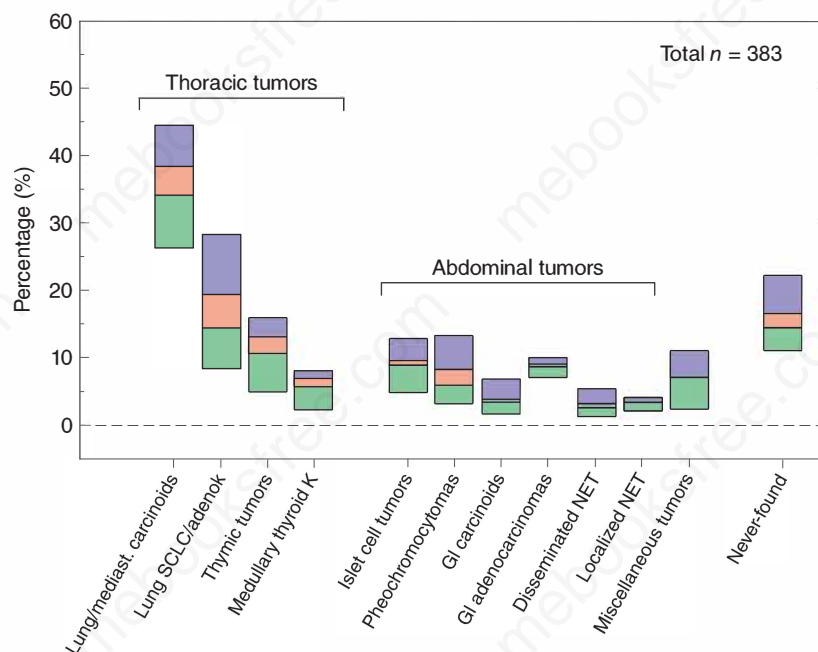
FIGURE 9–11 Prevalence of the most common tumors causing ectopic ACTH secretion. (K, carcinoma; NET, neuroendocrine tumor; SCLC, small cell lung carcinoma). (Reproduced with permission from Isidori AM, Lenzi A. Ectopic ACTH syndrome. *Arq Bras Endocrinol Metabol*. 2007 Nov;51(8):1217-1225.)

familial cause of adrenal-dependent Cushing syndrome. It is associated with myxomas (cardiac, cutaneous, and mammary), spotty skin pigmentation, endocrine overactivity, sexual precocity, acromegaly, and schwannomas—referred to as the Carney complex. Germline inactivating mutations in the regulatory subunit of protein kinase A have been shown to cause constitutive activation of a pathway critical for cortisol production.

3. **Bilateral macronodular adrenal hyperplasia (BMAH)**—Multiple distinct entities exist that can result in BMAH. These adrenal nodules may be under the control of abnormal or ectopic hormone receptors. Aberrant regulation of cortisol production and adrenal growth has been shown in some of these patients to be mediated by the abnormal adrenal expression of receptors for a variety of hormones. The best characterized appears to be the expression of glucose-dependent insulinotropic polypeptide (GIP), whose adrenal expression has resulted in the modulation of cortisol production after physiologic postprandial fluctuation of endogenous levels of GIP, causing a state of *food-dependent Cushing syndrome*. Other abnormal hormone receptors that have been described in association with endogenous hypercortisolism with this phenomenon include vasopressin, beta-adrenergic agonists, human chorionic gonadotropin-LH, serotonin, angiotensin II, and leptin (see also Chapter 21). The identification of an ectopically expressed adrenal receptor raises the possibility of new pharmacologic approaches for control of hypercortisolism by suppressing the endogenous ligands or by blocking the abnormally expressed receptor with specific antagonists.

Another recent discovery that has furthered our knowledge of adrenal tumor genetics is that of the role of the armadillo-repeat containing 5 (ARMC5) gene. ARMC5 is a putative tumor suppressor gene located on chromosome 16p11.2. An investigation of apparently sporadic cases of BMAH showed mutations in both alleles of ARMC5, one germline and one somatic. The inactivating mutations in ARMC5 appear to

decrease steroidogenesis *in vitro* and may explain why very large nodules are required to produce cortisol excess. Several studies have looked at relatives of patients with BMAH and found germline mutations in over half of the studied relatives. In these individuals without any symptoms of Cushing syndrome, adrenal nodular hyperplasia was frequently discovered. It has also been reported that ACTH may be produced in the adrenal glands of patients with BMAH. This intraadrenal ACTH production may be the reason why there is internodular hyperplasia in many patients with BMAH.

McCune-Albright syndrome (characterized by fibrous dysplasia of bone, café-au-lait skin pigmentation, precocious puberty, and endocrine hormone excess) is caused by activating mutations of $G_s\alpha$ that lead to constitutive steroidogenesis in adrenal nodules carrying the mutation.

Additionally, BMAH has been found to be associated with several other genetic syndromes including multiple endocrine neoplasia type 1, familial adenomatous polyposis, and hereditary leiomyomatosis and renal cell cancer syndrome.

Pathophysiology (Figures 9–10 and 9–12)

A. Cushing disease   In Cushing disease, ACTH hypersecretion is random and episodic and causes cortisol hypersecretion in the absence of the normal circadian rhythm. Feedback inhibition of ACTH (secreted from the pituitary adenoma) by physiologic levels of glucocorticoids is suppressed; thus, ACTH hypersecretion persists despite elevated cortisol secretion and results in chronic glucocorticoid excess. The episodic secretion of ACTH and cortisol results in variable plasma levels that may at times be within the normal range. However, demonstration of elevated late-night serum or salivary cortisol levels, lack of suppression of cortisol after dexamethasone, or elevation of urine

Cushing syndrome suspected
(consider endocrinologist consultation)

Exclude exogenous glucocorticoid exposure

Perform one of the following tests

24-h UFC (≥2 tests)   Overnight   Late-night salivary
                             1-mg DST   cortisol (≥2 tests)

Consider caveats for each test (see text)
Use 48-h, 2-mg DST in certain populations (see text)

ANY ABNORMAL RESULT   Normal (CS unlikely)

Exclude physiologic causes of hypercortisolism (Table 9–2)

Consult endocrinologist

Perform 1 or 2 other studies shown above

Suggest consider or repeating the abnormal study
Suggest Dex-CRH or midnight serum cortisol in
certain populations (see text)

Discrepant   ABNORMAL   Normal (CS unlikely)
(Suggest additional evaluation)

Cushing syndrome

FIGURE 9–12   The diagnosis of Cushing syndrome. (Reproduced with permission from Nieman LK, Biller BM, Findling JW, et al. The diagnosis of Cushing syndrome: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab.* 2008 May;93(5):1526-1540.)

free cortisol confirms cortisol hypersecretion (see sections on Laboratory Evaluation and Diagnosis of Cushing Syndrome). The overall increase in glucocorticoid secretion causes the clinical manifestations of Cushing syndrome. ACTH and **β-lipotropin** secretion are not usually elevated sufficiently to cause hyperpigmentation.

1. **Abnormalities of ACTH secretion**—Despite ACTH hypersecretion, stress responsiveness is absent; stimuli such as hypoglycemia or surgery fail to further elevate ACTH and cortisol secretion. This is probably due to suppression of hypothalamic function and CRH secretion by hypercortisolism, resulting in loss of hypothalamic control of ACTH secretion (see Chapter 4).
2. **Effect of cortisol excess**—Cortisol excess not only inhibits normal pituitary and hypothalamic function, affecting ACTH, TSH, GH, and gonadotropin release, but also results in all the systemic effects of glucocorticoid excess described in previous sections and in the section on Clinical Features later.
3. **Androgen excess**—Secretion of adrenal androgens is also increased in Cushing disease, and the degree of androgen excess parallels that of ACTH and cortisol. Thus, plasma levels of DHEA, DHEA sulfate, and androstenedione may be moderately elevated in Cushing disease; the peripheral conversion of

these hormones to testosterone and dihydrotestosterone leads to androgen excess. In women, this causes hirsutism, acne, and amenorrhea. In men with Cushing disease, cortisol suppression of LH secretion decreases testosterone secretion by the testis, resulting in decreased libido and impotence. The increased adrenal androgen secretion is insufficient to compensate for the decreased gonadal testosterone production.

B. Ectopic ACTH syndrome   Hypersecretion of ACTH and cortisol is usually greater in patients with ectopic ACTH syndrome than in those with Cushing disease. ACTH and cortisol hypersecretion is randomly episodic, and the levels are often greatly elevated. Usually, ACTH secretion by ectopic tumors is not subject to negative feedback control; that is, secretion of ACTH and cortisol is nonsuppressible with pharmacologic doses of glucocorticoids (see section on Diagnosis later).

Plasma levels, secretion rates, and urinary excretion of cortisol, the adrenal androgens, and DOC are often markedly elevated; despite this, the typical features of Cushing syndrome maybe absent, presumably because of rapid onset of hypercortisolism, anorexia, and other manifestations of the associated malignant disease. Features of mineralocorticoid excess (hypertension and

hypokalemia) are frequently present and have been attributed to increased secretion of DOC and the mineralocorticoid effects of cortisol. With ectopic CRH secretion, pituitary corticotroph hyperplasia and ACTH hypersecretion are observed along with resistance to negative feedback by cortisol.

C. Adrenal tumors

1. **Autonomous secretion**—Primary adrenal tumors, both adenomas and carcinomas, autonomously hypersecrete cortisol. Circulating plasma ACTH levels are suppressed, resulting in cortical atrophy of the uninvolved adrenal. Secretion is randomly episodic, and these tumors are typically unresponsive to manipulation of the hypothalamic-pituitary axis with pharmacologic agents such as dexamethasone and metyrapone.

2. **Adrenal adenomas**—Adrenal adenomas causing Cushing syndrome typically present solely with clinical manifestations of glucocorticoid excess, because they usually secrete only cortisol. Thus, the presence of androgen or mineralocorticoid excess should suggest that the tumor is an adrenocortical carcinoma.

3. **Adrenal carcinomas**—Adrenal carcinomas frequently hypersecrete multiple adrenocortical steroids and their precursors. Cortisol and androgens are the steroids most frequently secreted in excess; 11-deoxycortisol is often elevated, and there may be increased secretion of DOC, aldosterone, or estrogens. Plasma cortisol and urine free cortisol are often markedly increased; androgen excess is usually even greater than that of cortisol. Thus, high levels of plasma DHEA, DHEA sulfate, and of testosterone typically accompany the cortisol excess. Clinical manifestations of hypercortisolism are usually severe and rapidly progressive in these patients. In women, features of androgen excess are prominent; virilism may occasionally occur. Hypertension and hypokalemia are frequent and most commonly result from the mineralocorticoid effects of cortisol; less frequently, DOC and aldosterone hypersecretion also contribute.

Clinical Features (Table 9–13)

A. Symptoms and signs

1. **Obesity**—Obesity is the most common manifestation, and weight gain is usually the initial symptom. It is classically central, affecting mainly the face, neck, trunk, and abdomen, with relative sparing of the extremities. Generalized obesity with central accentuation is equally common, particularly in children.

   Accumulation of fat in the face leads to the typical facial rounding, *moon facies*, which is present in 75% of cases and is accompanied by facial plethora in most patients. Fat accumulation around the neck is prominent in the supraclavicular and dorsocervical fat pads.

   Obesity is absent in a handful of patients who do not gain weight; however, they usually have central redistribution of fat and a typical facial appearance.

2. **Skin changes**—Skin changes are frequent, and their presence should arouse a suspicion of cortisol excess. Atrophy of the epidermis and its underlying connective tissue leads to thinning (a transparent appearance of the skin) and facial plethora. Easy bruisability following minimal trauma is present in about 40%. Striae occur in 50% but are very unusual in patients over 40 years of age. These are typically red to purple, depressed below the skin surface secondary to loss of

TABLE 9–13 Clinical features of Cushing syndrome (% prevalence).

General
Obesity 90%
Hypertension 85%

Skin
Plethora 70%
Hirsutism 75%
Striae 50%
Acne 35%
Bruising 35%

Musculoskeletal
Osteopenia 80%
Weakness 65%

Neuropsychiatric 85%
Emotional lability
Euphoria
Depression
Psychosis

Gonadal Dysfunction
Menstrual disorders 70%
Impotence, decreased libido 85%

Metabolic
Glucose intolerance 75%
Diabetes 20%
Hyperlipidemia 70%
Polyuria 30%
Kidney stones 15%

underlying connective tissue, and wider (not infrequently >1-2 cm) than the pinkish-white striae that may occur with pregnancy or rapid weight gain. These striae are most commonly abdominal but may also occur over the breasts, hips, buttocks, thighs, and axillae.

Acne presenting as pustular or papular lesions may result from glucocorticoid excess or hyperandrogenism.

Minor wounds and abrasions may heal slowly, and surgical incisions sometimes undergo dehiscence.

Mucocutaneous fungal infections are frequent, including tinea versicolor, involvement of the nails (onychomycosis), and oral candidiasis.

Hyperpigmentation of the skin is rare in Cushing disease or adrenal tumors but is common in ectopic ACTH syndrome.

3. **Hirsutism**—Hirsutism is present in about 80% of female patients owing to hypersecretion of adrenal androgens. Facial hirsutism is most common, but increased hair growth may also occur over the abdomen, breasts, chest, and upper thighs. Acne and seborrhea usually accompany hirsutism. Virilism is unusual except in cases of adrenal carcinoma, in which it occurs in about 20% of patients.

4. **Hypertension**—Hypertension is a classic feature of spontaneous Cushing syndrome; it is present in about 75% of cases, and the diastolic blood pressure is greater than 100 mm Hg in over 50%. Hypertension and its complications contribute greatly to the morbidity and mortality rates in spontaneous Cushing syndrome.

5. **Gonadal dysfunction**—This is very common as a result of elevated androgens (in females) and cortisol (in males and to a lesser extent in females). Amenorrhea occurs in 75% of premenopausal women and is usually accompanied by

infertility. Decreased libido is frequent in males, and some have decreased body hair and soft testes.

6. **Central nervous system and psychologic disturbances—**Psychologic disturbances occur in the majority of patients. Mild symptoms consist of emotional lability and increased irritability. Anxiety, depression, poor concentration, and poor memory may also be present. Euphoria is frequent, and occasionally patients manifest overtly manic behavior. Sleep disorders are present in most patients, with either insomnia or early morning awakening.

   Severe psychologic disorders occur in a few patients and include severe depression, psychosis with delusions or hallucinations, and paranoia. Some patients have committed suicide. Loss of brain volume that is at least partially reversible following correction of hypercortisolism has been observed in several reports.

7. **Muscle weakness—**This occurs in about 60% of cases; it is more often proximal and is usually most prominent in the lower extremities. Hypercortisolism is associated with both low fat-free muscle mass and low total body protein.

8. **Osteoporosis—**Owing to the profound effects of glucocorticoids on the skeleton, patients with Cushing syndrome frequently have evidence of significant osteopenia and osteoporosis. Patients may present with multiple fragility fractures—typically of the feet, ribs, or vertebrae. Back pain may be the initial complaint. Compression fractures of the spine are demonstrated radiographically in 15% to 20% of patients. In fact, unexplained osteopenia in any young or middle-aged adult should always prompt an evaluation for Cushing syndrome, even in the absence of any other signs or symptoms of cortisol excess. Although avascular necrosis of bone has been associated with exogenous glucocorticoid administration, the problem is rarely observed in patients with endogenous hypercortisolism.

9. **Renal calculi—**Calculi secondary to glucocorticoid-induced hypercalciuria occur in approximately 15% of patients, and renal colic may occasionally be a presenting complaint.

10. **Thirst and polyuria—**Polyuria is rarely due to overt hyperglycemia but is usually caused by glucocorticoid-induced inhibition of vasopressin (antidiuretic hormone or ADH) secretion and the direct enhancement of renal free water clearance by cortisol.

11. **Growth deceleration—**In children, nearly all (>95%) show decreasing linear growth. Weight gain accompanying decreasing linear growth in a child should prompt an evaluation for Cushing syndrome.

B. Laboratory findings   Routine laboratory examinations are described here. Specific tests to establish the diagnosis of Cushing syndrome are discussed in the section on Diagnosis.

High normal hemoglobin, hematocrit, and red cell counts are usual; polycythemia is rare. The total white count is usually normal; however, both the percentage of lymphocytes and the total lymphocyte count may be subnormal. Eosinophils are also depressed, and a total eosinophil count less than $100/\mu L$ is present in most patients. Serum electrolytes, with rare exceptions, are normal in Cushing disease; however, hypokalemic alkalosis occurs when there is marked steroid hypersecretion with the ectopic ACTH syndrome or adrenocortical carcinoma.

Although fasting hyperglycemia occurs in only 10% to 15% of patients, glucose intolerance is a relatively common finding and occurs in 60% of patients. Most patients have secondary hyperinsulinemia and abnormal glucose tolerance tests.

Serum calcium is normal; serum phosphorus is low normal or slightly depressed. Hypercalciuria is present in 40% of cases.

C. Imaging studies   Routine radiographs may reveal cardiomegaly, due to hypertensive or atherosclerotic heart disease, or mediastinal widening, due to central fat accumulation. Vertebral compression fractures, rib fractures, and renal calculi may be present.

D. Electrocardiographic findings   Hypertensive, ischemic, and electrolyte-induced changes may be present on the electrocardiogram.

Features Suggesting a Specific Cause

A. Cushing disease   Cushing disease typifies the classic clinical picture: female predominance, onset generally between ages 20 and 40, and a slow progression over several years. Hyperpigmentation and hypokalemic alkalosis are rare; androgenic manifestations are limited to acne and hirsutism. In general, secretion of cortisol and adrenal androgens is only moderately increased.

B. Ectopic ACTH syndrome (carcinoma)   In contrast, this syndrome occurs predominantly in males, with the highest incidence between ages 40 and 60. The clinical manifestations of hypercortisolism are frequently limited to weakness, hypertension, and glucose intolerance; the primary tumor is usually apparent. Hyperpigmentation, hypokalemia, and alkalosis are common, as are weight loss and anemia. The hypercortisolism is of rapid onset, and steroid hypersecretion is frequently severe, with equally elevated levels of glucocorticoids, androgens, and DOC.

C. Ectopic ACTH syndrome (benign tumor)   A minority of patients with ectopic ACTH syndrome due to more benign tumors, especially bronchial carcinoids, present a more slowly progressive course, with typical features of Cushing syndrome. These patients may be clinically identical to those with pituitary-dependent Cushing disease, and the responsible tumor may not be apparent. Hyperpigmentation, hypokalemic alkalosis, and anemia are variably present. Further confusion may arise, since a number of these patients with occult ectopic tumors may have ACTH and steroid dynamics typical of Cushing disease (see later).

D. Adrenal adenomas   The clinical picture in patients with adrenal adenomas is usually that of glucocorticoid excess alone, and androgenic effects such as hirsutism are absent. Onset is gradual, and hypercortisolism is mild to moderate. Plasma androgens are usually in the low normal or subnormal range.

E. Adrenal carcinomas   In general, adrenal carcinomas have a rapid onset of the clinical features of excessive glucocorticoid, androgen, and mineralocorticoid secretion and are rapidly progressive. Marked elevations of both cortisol and androgens are usual; hypokalemia is common, as are abdominal pain, palpable masses, and hepatic and pulmonary metastases.

Diagnosis

The clinical suspicion of Cushing syndrome must be confirmed with biochemical studies. Initially, a general assessment of the patient regarding the presence of other illnesses, drugs, alcohol, or psychiatric problems must be done since these factors may confound the evaluation. In the majority of cases, the biochemical diagnosis of Cushing syndrome can be easily performed in the ambulatory setting (see Figure 9–12).

A. Dexamethasone suppression test   The overnight 1-mg dexamethasone suppression test is a valuable screening test in patients with suspected hypercortisolism. This study employs the administration of 1 mg of dexamethasone at bedtime (11:00 PM), with determination of a plasma cortisol early the following morning. Normal subjects should suppress plasma cortisol to less than 1.8 μg/dL (50 nmol/L) following an overnight 1-mg test. Although a level of less than 5 μg/dL has been used in the past, several false-negative studies have been discovered using this test criterion. False-negative results may occur in some patients with mild hypercortisolism and in those with intermittent hypercortisolism. False-positive results with the overnight 1-mg dexamethasone suppression test may be caused by drugs that accelerate dexamethasone metabolism (phenytoin, phenobarbital, rifampin) or by drugs that increase CBG and resultant total cortisol levels (estrogens). False-positive results also occur in patients with renal failure, depression, alcoholism, or in patients undergoing a stressful event or serious illness.

B. Urine free cortisol   Another study used in the diagnosis of Cushing syndrome is the determination of urine free cortisol measured by HPLC or LC/MS/MS in a 24-hour urine collection. These methods are highly accurate and specific for the measurement of cortisol. Commonly used drugs and medications do not interfere; however, carbamazepine and fenofibrate can cause falsely elevated results with HPLC since they elute with cortisol. Urinary free cortisol is usually less than 50 μg/24 h (<135 nmol/24 h) when measured by HPLC or LC/MS/MS. The major liability of urine free cortisol is its poor sensitivity for detecting patients with mild to moderate hypercortisolism (sensitivity <75%). Consequently, this modality should not be used alone to exclude Cushing syndrome.

C. Late-night serum and salivary cortisol   The absence of diurnal cortisol rhythm is a hallmark of Cushing syndrome. Normally, cortisol is secreted episodically with a diurnal rhythm paralleling the secretion of ACTH. Levels are usually highest early in the morning and decrease gradually throughout the day, reaching the nadir in the late evening between 11:00 PM and midnight. Several studies have demonstrated that an elevated midnight plasma cortisol level (>5.2-7.0 μg/dL [140-190 nmol/L]) is highly accurate in differentiating patients with Cushing syndrome from normal subjects and from patients with pseudo-Cushing conditions such as depression or alcoholism.

Because obtaining such plasma cortisol measurements is impractical on an ambulatory basis, the measurement of salivary cortisol provides a simple and more convenient means of probing nighttime cortisol secretion. Saliva can easily be sampled at home by the patient using a variety of techniques. Reference ranges for late-night salivary cortisol concentrations are dependent on the assay employed; however, using a radioimmunoassay, normal subjects should have values less than 0.15 μg/dL (4 nmol/L).

Problems in Diagnosis

It has become increasingly appreciated that many common physiologic situations and medical disorders can also cause sustained or intermittent hypercortisolism and may mimic—clinically and biochemically—the neoplastic causes of Cushing syndrome with which we are so familiar (pituitary/nonpituitary ACTH-secreting tumors and adrenal neoplasia). These disorders have been characterized as "pseudo-Cushing syndrome." Of course, the clinical features of Cushing syndrome such as obesity, hypertension, diabetes, myopathy, osteoporosis, and neuropsychiatric problems may be evident in some patients with chronic physiologic/nonneoplastic hypercortisolism (eg, depression, chronic alcoholism, and chronic kidney disease) and be indistinguishable from those with pathologic Cushing syndrome.

These conditions may have biochemical features of Cushing syndrome, including elevations of urine free cortisol, disruptions in the normal diurnal pattern of cortisol secretion, and lack of suppression of cortisol after the overnight 1-mg dexamethasone suppression test. Although the history and physical examination may provide specific clues to the appropriate diagnosis, definitive biochemical confirmation may be difficult and may require repeated testing.

The ACTH and cortisol responses to DDAVP have been used to distinguish patients with pathologic/neoplastic Cushing syndrome from those with physiologic/non-neoplastic hypercortisolism. Since corticotroph adenomas harbor vasopressin receptors, patients with Cushing disease have an exaggerated ACTH response to desmopressin (peak response >70 pg/mL [>15.4 pmol/L] or an increase >27 pg/mL [>5.9 pmol/L]), while those with alcohol-induced Cushing syndrome (and other nonneoplastic hypercortisolemic states) have a blunted ACTH response. The cortisol responses have been less valuable in this differentiation.

Another study to distinguish mild Cushing syndrome from pseudo-Cushing conditions is the use of dexamethasone suppression followed by CRH stimulation. This test takes advantage of differential response of patients with Cushing syndrome and pseudo-Cushing syndrome to both dexamethasone and CRH by combining these tests to provide greater accuracy in the diagnosis. This study involves the administration of dexamethasone, 0.5 mg every 6 hours for eight doses, followed immediately by a CRH stimulation test, starting 2 hours after the completion

of the low-dose dexamethasone suppression. A plasma cortisol concentration greater than 1.4 μg/dL (38.6 nmol/L) measured 15 minutes after administration of CRH correctly identifies the majority of patients with Cushing syndrome. After administration of high-dose dexamethasone, patients with pseudo-Cushing syndrome do not have a rise in cortisol or ACTH in response to CRH. However, it is now appreciated that some patients with pseudo-Cushing syndrome (eg, anorexia nervosa) may have abnormal dexamethasone-CRH suppression tests.

Differential Diagnosis

The differential diagnosis of Cushing syndrome can be difficult and should always be performed in consultation with an endocrinologist. The introduction of several technologic advances, including a specific and sensitive IRMA or ICMA for ACTH, CRH stimulation test, inferior petrosal sinus sampling (IPSS), CT and magnetic resonance imaging (MRI) of the pituitary and adrenal glands have all provided means for an accurate differential diagnosis.

A. Plasma ACTH Initially, the differential diagnosis for Cushing syndrome must distinguish between ACTH-dependent Cushing syndrome (pituitary or nonpituitary ACTH-secreting neoplasm) and ACTH-independent hypercortisolism. The best way to distinguish these forms of Cushing syndrome is measurement of plasma ACTH by IRMA or ICMA. The development of this sensitive and specific test has made it possible to reliably identify patients with ACTH-independent Cushing syndrome. The ACTH level is generally less than 10 pg/mL (2.2 pmol/L) and exhibits a blunted response to CRH (peak response

<10 pg/mL [2.2 pmol/L]) in patients with cortisol-producing adrenal neoplasms, autonomous bilateral adrenal cortical hyperplasia, and factitious Cushing syndrome (Figure 9–13). Patients with ACTH-secreting neoplasms usually have plasma ACTH levels greater than 20 pg/mL (4.4 pmol/L) and frequently greater than 52 pg/mL (11.5 pmol/L). The major challenge in the differential diagnosis of ACTH-dependent Cushing syndrome is identifying the source of the ACTH-secreting tumor. The vast majority of these patients (90%) have a pituitary tumor, whereas the others harbor a nonpituitary neoplasm. Diagnostic studies needed to differentiate these two entities must yield nearly perfect sensitivity, specificity, and accuracy. Although plasma ACTH levels are usually higher in patients with ectopic ACTH than those with pituitary ACTH-dependent Cushing syndrome, there is considerable overlap between these two entities. Many of the ectopic ACTH-secreting tumors are radiologically occult at the time of presentation and may not become clinically apparent for many years after the initial diagnosis. However, an enhanced ACTH response for CRH administration is more frequently found in pituitary Cushing syndrome compared with ectopic ACTH syndrome.

B. Pituitary MRI When ACTH-dependent Cushing syndrome is present, MRI of the pituitary gland with gadolinium enhancement should be performed and will identify an adenoma in at least 50% to 60% of the patients. If the patient has classic clinical laboratory findings of ACTH-dependent hypercortisolemia and an unequivocal pituitary lesion on MRI, the likelihood of Cushing disease is 98% to 99%. However, it must be emphasized that approximately 10% of the population in the age group from 20 to 50 years will have incidental tumors of the

ACTH-IRMA measurements in pituitary-adrenal disorders
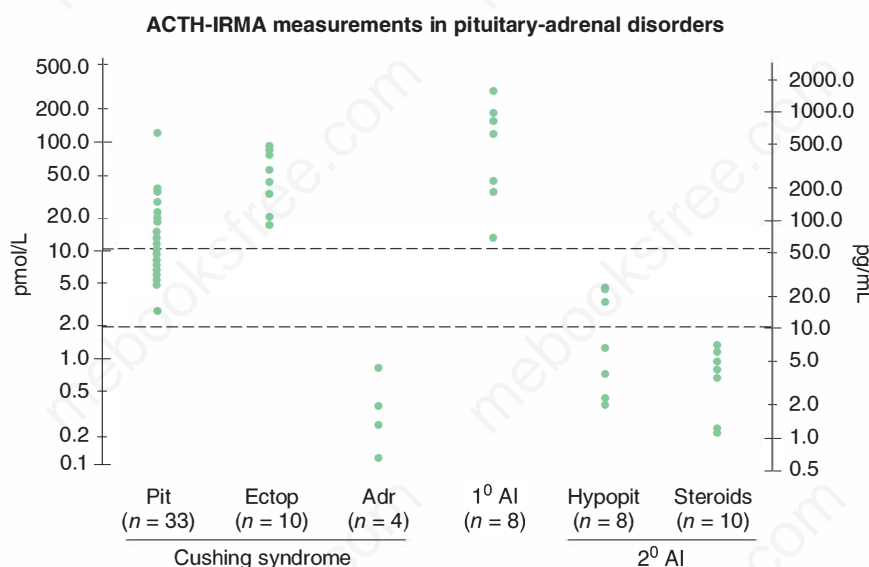
FIGURE 9–13 Plasma ACTH-IRMA (pmol/L or pg/mL) of patients with pituitary-adrenal disorders. Dashed horizontal lines indicate reference range. CD = Cushing disease, EAS = Ectopic ACTH syndrome, ACS = Adrenal ACTH-independent Cushing syndrome, 1⁰ AI = primary adrenal insufficiency, 2⁰ AI = Secondary adrenal insufficiency, Pit = Secondary adrenal insufficiency due to pituitary disease, Steroids = Secondary adrenal insufficiency due to exogenous steroid use.

pituitary demonstrable by MRI. Therefore, some patients with ectopic ACTH syndrome will have radiographic evidence of a pituitary lesion.

C. High-Dose dexamethasone suppression   Traditionally, the high-dose dexamethasone suppression test has been utilized in the differential diagnosis of Cushing syndrome. However, the diagnostic accuracy of this procedure is only 70% to 80%, which is actually less than the pretest probability of Cushing disease—on average about 90%. As a result, this study is no longer recommended.

D. Inferior petrosal sinus sampling (IPSS)   The most definitive means of accurately distinguishing pituitary from nonpituitary ACTH-dependent Cushing syndrome is the use of bilateral simultaneous IPSS with CRH stimulation, and this procedure is the next step in the evaluation of patients with ACTH-dependent Cushing syndrome when MRI does not reveal a definite adenoma. This study takes advantage of the means by which pituitary hormones reach the systemic circulation. Blood leaves the anterior lobe of the pituitary and drains into the cavernous sinuses, which then empty into the inferior petrosal sinuses and subsequently into the jugular bulb and vein. Simultaneous inferior petrosal sinus and peripheral ACTH measurement before and after CRH stimulation can reliably confirm the presence or absence of an ACTH-secreting pituitary tumor. An inferior petrosal sinus to peripheral (IPS-P) ACTH ratio greater than 2.0 prior to CRH and greater than 3.0 after CRH is consistent with a pituitary ACTH-secreting tumor, and an IPSP ACTH ratio less than 1.8 supports the diagnosis of ectopic ACTH. Anatomical variants can cause false-negative studies; however, simultaneous measurement of PRL can be used to construct a PRL-normalized ACTH ratio that can eliminate false negatives caused by anatomical variants.

Bilateral IPSS with CRH stimulation does require a skilled interventional radiologist. However, in experienced hands, the procedure has yielded a diagnostic accuracy approaching 100% in identifying the source of ACTH-dependent Cushing syndrome.

E. Occult ectopic ACTH   If the IPSS study is consistent with a nonpituitary ACTH-secreting tumor, a search for an occult ectopic ACTH-secreting tumor is needed. Because the majority of these lesions are in the thorax, high-resolution CT of the chest may be useful; MRI of the chest may have even better sensitivity in finding these lesions, which are usually small bronchial carcinoid tumors. Unfortunately, utilization of a radiolabeled somatostatin analog scan (octreotide acetate scintigraphy) has met with only mixed results in localizing these tumors.

F. Adrenal localizing procedures   CT (Figure 9–14) and MRI are used to define adrenal lesions. Their primary use is to localize adrenal tumors in patients with ACTH-independent Cushing syndrome. Most adenomas exceed 2 cm in diameter
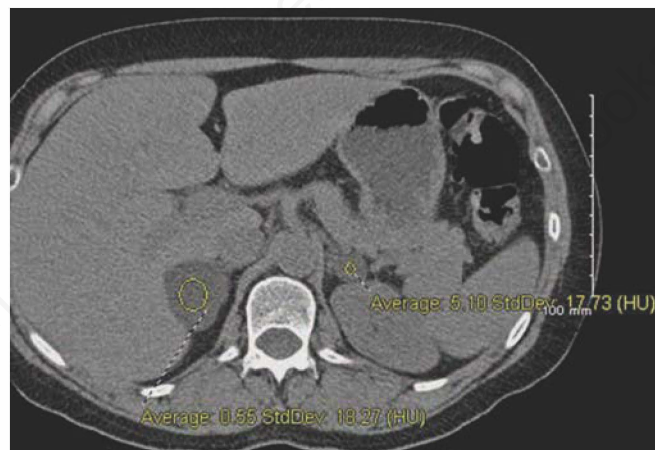
FIGURE 9–14   Adrenal CT scan in Cushing syndrome. The patient has family history of macronodular adrenal disease with associated Cushing syndrome. Both adrenal glands visualized in this CT have nodular enlargement with the right adrenal gland showing significant macronodular hyperplasia. The adrenal nodules seen here have low CT attenuation, indicative of non-malignant lesions.

and have a low density on CT; carcinomas are usually much larger with irregular borders and have high density and low contrast washout on CT.

TREATMENT

A. Cushing disease   The aim of treatment of Cushing syndrome is to remove or destroy the causative lesion and thus correct hypersecretion of adrenal hormones without inducing pituitary or adrenal damage, which requires permanent replacement therapy for hormone deficiencies.

The treatment of choice for Cushing disease is transsphenoidal pituitary microsurgery with adenomectomy. Complete removal of the ACTH-secreting pituitary lesion resolves ACTH hypersecretion and generally induces transient adrenal insufficiency. This is due to suppression of normal pituitary corticotrophs from cortisol excess. In most cases, the adrenal insufficiency resolves within several months, and normal cortisol dynamics are restored. If complete resection of the causative lesion is not possible, radiation therapy to the pituitary lesion or medical therapy to lower cortisol levels can be implemented. In some cases bilateral adrenalectomy is needed to control the hypercortisolism. These methods are discussed in greater detail in Chapter 4.

B. Ectopic ACTH syndrome   Cure of ectopic ACTH syndrome is usually possible only in cases involving the more benign tumors such as bronchial or thymic carcinoids or pheochromocytomas. Treatment is made difficult by the presence of metastatic tumors and accompanying severe hypercortisolism. If therapy directed to the primary tumor is unsuccessful, other means (such as bilateral adrenalectomy, glucocorticoid inhibitors, or glucocorticoid receptor antagonists) must be used to correct the steroid-excess state.

Severe hypokalemia may require potassium replacement in large doses and spironolactone to block mineralocorticoid effects.

C. Adrenal tumors

1. **Adrenal adenomas**—Patients with adrenal adenomas are successfully treated by unilateral adrenalectomy, and the outlook is excellent. Laparoscopic adrenalectomy has become widely used in most patients with benign adrenal tumors and has significantly reduced the duration of the hospital stay. Because the hypothalamic-pituitary axis and the contralateral adrenal are suppressed by prolonged cortisol secretion, these patients often have postoperative adrenal insufficiency and require glucocorticoid therapy both during and following surgery until the remaining adrenal recovers.
2. **Adrenal carcinomas**—Therapy in cases of adrenocortical carcinoma is less satisfactory, because the tumor has frequently already metastasized (usually to the retroperitoneum, liver, and lungs) by the time the diagnosis is made.
   a. **Operative treatment**—Surgical cure depends on the size and extent of the tumor, and cure is not possible in metastatic or locally advanced disease. Persisting nonsuppressible steroid secretion in the immediate postoperative period indicates residual or metastatic tumor.
   b. **Medical treatment**—Mitotane is the drug of choice. The dosage is 6 to 12 g/d orally in three or four divided doses. The dose must be reduced because of side-effects in 80% of patients (diarrhea, nausea and vomiting, depression, somnolence). About 70% of patients achieve a reduction of steroid secretion, but only 35% achieve a reduction in tumor size.

Ketoconazole and metyrapone (singly or in combination) are useful in controlling steroid hypersecretion in patients who do not respond to mitotane.

D. Bilateral macrodular adrenal hyperplasia

Bilateral adrenalectomy is curative in these patients, but the decision to perform bilateral adrenalectomy must be weighed against the resultant adrenal insufficiency and its inherent morbidity. Some evidence exists that unilateral adrenalectomy of the largest adrenal gland can provide control of the hypercortisolism for at least some time before removal of the other adrenal gland is needed.

PROGNOSIS

A. Cushing syndrome
Untreated Cushing syndrome is frequently fatal, and death may be due to the underlying tumor itself, as in the ectopic ACTH syndrome and adrenal carcinoma. However, in most cases, death is the consequence of sustained hypercortisolism and its complications, including hypertension, cardiovascular disease, stroke, thromboembolism, and susceptibility to infection.

B. Cushing disease
In contrast to patients with untreated Cushing syndrome, individuals with Cushing disease who achieve normalization of cortisol levels as result of therapy have a standard mortality ratio (SMR) similar to age-matched populations. However, patients with persistent hypercortisolism despite treatment continue to have an increased SMR (3.8- to 5.0-fold), when compared to the general population.

With current pituitary microsurgery, the majority of patients with Cushing disease can be treated successfully with remission rates of 65% to 90% in patients with tumors less than 1 cm and approximately 65% in patients with tumors more than 1 cm. Even with initial remission, up to 25% of patients with small tumors may have recurrence at 5 to 10 years after therapy. Unfortunately, reduction in health-related quality of life, especially related to psychological functioning, and metabolic derangements may persist after successful surgery.

C. Adrenal tumors
The prognosis in adrenal adenomas is excellent. In adrenal carcinoma, the prognosis is almost universally poor, and the median survival from the date of onset of symptoms is about 4 years.

D. Ectopic ACTH syndrome
Prognosis is also poor in patients with ectopic ACTH syndrome due to the nature of the malignancy producing the hormone, and in these patients with severe hypercortisolism, survival is frequently only days to weeks. Some patients respond to tumor resection or chemotherapy. The prognosis is better in patients with benign tumors producing the ectopic ACTH syndrome.

HIRSUTISM AND VIRILISM

Excessive adrenal or ovarian secretion of androgens or excessive conversion of weak androgens to testosterone and dihydrotestosterone in peripheral tissues leads to hirsutism and virilism (see Chapter 13). As previously discussed, the adrenal secretory products DHEA, DHEA sulfate, and androstenedione are weak androgens; however, the peripheral conversion to testosterone and dihydrotestosterone can result in a state of androgen excess.

Excessive androgen production is seen in both adrenal and ovarian disorders. Adrenal causes include Cushing syndrome, adrenal carcinoma, and congenital adrenal hyperplasia (see previous sections and Chapter 14). Mild adult-onset cases of congenital adrenal enzyme deficiencies have been described; these appear to be relatively uncommon. Biochemical diagnosis of late-onset 21-hydroxylase deficiency is best achieved by measurement of the 17-hydroxyprogesterone response to ACTH. Ovarian causes are discussed in Chapter 13.

In children, androgen excess is usually due to premature adrenarche, congenital adrenal hyperplasia, or adrenal carcinoma. In women, hirsutism accompanied by amenorrhea, infertility, and ovarian enlargement is typical of the polycystic ovary syndrome, whereas in Cushing syndrome hirsutism is accompanied by features of cortisol excess. Late-onset 21-hydroxylase deficiency is accompanied by elevated levels of plasma 17-hydroxyprogesterone, especially following ACTH administration. Virilism and severe androgen excess in adults are usually due to androgen-secreting adrenal or ovarian tumors; virilism is unusual in the polycystic ovary syndrome and rare in Cushing disease. In the absence of these syndromes, hirsutism in women

is usually idiopathic or due to milder forms of polycystic ovary syndrome. Exogenous androgen administration (eg, DHEA) should also be considered.

The diagnosis and therapy of hirsutism are discussed in Chapter 13.

INCIDENTAL ADRENAL MASS

The incidental adrenal mass has become a common diagnostic problem, because approximately 2% to 4% of patients undergoing CT studies of the abdomen are found to have focal enlargement of the adrenal gland. Adrenal masses in an adult may represent functional or nonfunctional cortical adenomas or carcinoma, pheochromocytomas, cysts, myelolipomas, or metastasis from other tumors. Congenital adrenal hyperplasia may also uncommonly present as a focal enlargement of the adrenal gland, and adrenal hemorrhage also causes enlargement.

Investigation of an incidental adrenal mass should focus on two distinct areas: (1) exclusion of malignancy; and (2) exclusion of hormonal excess.

Exclusion of Malignancy

When trying to determine if an adrenal mass is malignant, lesion size and appearance are helpful features, as most adrenal carcinomas and metastatic lesions are large (>4 cm) and have irregular or invasive borders. The most definitive features used to exclude malignancy are CT density and contrast washout. Lesions with low density (<10 Hounsfield units) on unenhanced CT scans are almost certainly benign, but some benign lesions have a density more than 10 Hounsfield units. In cases where the lesion does not have a low density, administration of contrast with delayed imaging is very helpful. Benign lesions have a contrast washout of more than 50%, whereas malignant lesions have a low amount of contrast washout.

MRI of the adrenal gland is usually not necessary but may be helpful in selected patients. Typically, malignancies and pheochromocytomas tend to have bright signal intensity with T2-weighted images. Percutaneous adrenal biopsy really has no demonstrated efficacy in patients with adrenal masses and no history of a malignancy. Percutaneous adrenal biopsy should be reserved for patients in whom the presence or absence of adrenal metastases may alter the therapy or prognosis of the patient. Pheochromocytoma should always be excluded before biopsy is undertaken.

Endocrine Evaluation

The appropriate biochemical evaluation of an incidental adrenal mass should include testing for pheochromocytoma, cortisol excess, and in patients with hypertension or hypokalemia, primary aldosteronism. To exclude pheochromocytoma, all patients with incidental adrenal masses should have a measurement of fractionated urinary or plasma metanephrines. Cortisol excess should also be evaluated in all patients. It appears that the overnight (1-mg) dexamethasone suppression test is the most sensitive test for cortisol excess in patients with mild adrenal-dependent cortisol excess.

A serum potassium and plasma aldosterone concentration-PRA ratio should be determined to evaluate for primary aldosteronism in individuals with hypertension or hypokalemia.

Repeat CT imaging in 6 to 12 months is also recommended in most cases. Demonstrating a lack of growth in a mass helps to further exclude neoplastic disease. Hormonal abnormalities may develop over time in nodules that were previously non-secretory, and follow-up testing has been recommended by some, depending on the clinical context.

Cortisol-Producing Adenoma

The most common functioning lesion in patients with an incidentally discovered adrenal mass appears to be an adenoma with autonomous secretion of cortisol. Traditional significant cortisol excess occurs in 6% to 9% of patients with incidental adrenal masses. Lesions with autonomous cortisol hypersecretion generally range in size from 2 to 5 cm in diameter. It is now well recognized that adrenal adenomas may secrete small but excessive amounts of cortisol. Although these lesions do not cause typical Cushing syndrome features, over time, the low level of cortisol excess is sufficient to contribute to low bone density, metabolic syndrome, and hypertension. Patients with prolonged, excessive low levels of autonomous cortisol secretion have been described as having mild cortisol excess or *subclinical* Cushing syndrome. The complete natural history of this autonomous cortisol secretion is unknown, but recent studies have shown an increase in cardiovascular morbidity and mortality in these patients. Studies have shown that surgical removal of adrenal adenomas in patients with even mild cortisol excess may result in improvement in their metabolic profile (ie, hypertension, obesity, diabetes) and a reduction in the incidence of future vertebral compression fractures. Consequently, adrenalectomy should be considered in patients with subclinical Cushing syndrome, especially in patients with clinical problems that might potentially be aggravated by glucocorticoid excess.

These patients are best identified by their failure to suppress cortisol to less than 1.8 μg/dL (50 nmol/L) following an overnight 1-mg dexamethasone suppression test. In addition, the basal levels of ACTH in these patients are subnormal or frankly suppressed. DHEA sulfate levels are also usually decreased in patients with cortisol-secreting adenomas, likely reflecting ACTH suppression. The autonomous cortisol secretion by the tumor may result in blunting of diurnal variation and in some cases, but not all, elevation of urine free cortisol levels. The low plasma ACTH level manifests in a blunted response to CRH administration. Removal of these silent adrenocortical adenomas may be followed by clinically significant secondary adrenal insufficiency.

Pheochromocytoma

Pheochromocytoma is a potentially life-threatening tumor that may present as an incidental adrenal mass. Surprisingly, pheochromocytoma may account for as many as 2% to 3% of incidental adrenal lesions. Many of these patients have hypertension and symptoms associated with catecholamine excess such as headache, diaphoresis, palpitations, or nervousness (see Chapter 11).

Aldosterone-Producing Adenoma

Although aldosterone-producing adenomas are more common than either pheochromocytomas or cortisol-producing adenomas, they actually represent a very unusual cause of an incidentally discovered adrenal mass. This appears to be due to the fact that aldosterone-producing adenomas are usually small and frequently missed with CT imaging of the adrenal gland. Because most of these patients have hypertension, this diagnosis needs to be considered only in patients with hypertension. The presence of hypokalemia should arouse suspicion of this diagnosis, but hypokalemia is not invariably present. It is usually present in patients with aldosterone-producing adenomas greater than 3 cm. Hyperaldosteronism can be excluded by the measurement of aldosterone and PRA. If the aldosterone (ng/dL)-PRA (ng/mL/h) ratio is less than 20 to 30 and plasma aldosterone is less than 14 ng/dL, an aldosterone-producing adenoma is excluded.

GLUCOCORTICOID THERAPY FOR NONENDOCRINE DISORDERS

Principles

Glucocorticoids have been used for their anti-inflammatory and immunosuppressive activity in treatment of a wide variety of disorders. These include rheumatologic disorders (eg, rheumatoid arthritis and systemic lupus erythematosus), pulmonary diseases (eg, asthma), renal disease (eg, glomerulonephritis), and many others. Because of their side-effects, glucocorticoids should be used in the minimum effective dose and for the shortest possible duration of therapy.

Synthetic Glucocorticoids

Steroid compounds have been synthesized, taking advantage of chemical alterations to the steroid nucleus that enhances glucocorticoid activity relative to mineralocorticoid activity. For example, prednisone has a double bond between positions 1 and 2 of cortisol and an 11-keto group instead of a hydroxyl group. It has three to five times more glucocorticoid activity than cortisol and relatively little mineralocorticoid activity. It must be converted to prednisolone by reduction of the 11-keto group to a hydroxyl group in order to be biologically active, a process that may be reduced in the presence of liver disease. Dexamethasone has the same additional double bond, a fluoro atom in the 9α position and a 16α-methyl group. This results in 10 to 20 times the glucocorticoid activity of cortisol and negligible mineralocorticoid activity. Many other compounds have been synthesized. Although most synthetic glucocorticoids exhibit little binding to CBG, their plasma half-lives are longer than that of cortisol.

Modes of Administration

Glucocorticoids may be administered parenterally, orally, topically, and by inhalation. Absorption rates from intramuscular and intra-articular sites depend on the particular glucocorticoid and its formulation. Transdermal absorption also depends on the severity of the inflammatory disorder, the area of the body to which the drug is applied, the presence of vehicles that enhance absorption (eg, urea), and the use of an occlusive dressing. Inhaled glucocorticoids vary in their bioavailability; the technique of administration (eg, use of spacers) also affects the amount of drug delivered to the lungs.

Side-Effects

In general, the severity of the side-effects is a function of dose and duration of therapy, but there is marked individual variation.

A. Hypothalamic-pituitary-adrenal axis suppression Glucocorticoids suppress CRH and ACTH secretion (negative feedback). Suppression of the HPA axis may occur at any dose of glucocorticoid, but generally with doses of prednisone greater than 5 mg/d. It is difficult, however, to predict the development or degree of suppression in any given individual. In general, patients who develop clinical features of Cushing syndrome or who have received glucocorticoids equivalent to 10 to 20 mg of prednisone per day for 3 weeks or more should be assumed to have clinically significant HPA axis suppression. Patients treated with alternate-day steroid regimens exhibit less suppression than those who receive steroids daily.

B. Cushing syndrome Glucocorticoid administration results in the development of Cushingoid features. Of special concern is steroid-induced osteoporosis, particularly in patients for whom a long course of steroid therapy is anticipated. The severity of systemic effects of inhaled glucocorticoids varies among different preparations. However, they are associated with both local effects (dysphonia and oral candidiasis) and systemic effects, especially glaucoma, cataracts, osteoporosis, and growth retardation in children. Protease inhibitors are now well recognized to decrease the metabolism of exogenous glucocorticoids. When the two are used in combination, the resultant increase in circulating glucocorticoid can lead to exogenous Cushing syndrome.

C. Steroid withdrawal Because of their adverse effects, glucocorticoids must be tapered downward as the clinical situation permits. Tapering regimens are essentially empirical. Factors that may limit the ability to taper the dose down to physiologic replacement levels include recrudescence of disease and steroid withdrawal syndrome. The latter appears in a variety of patterns. Patients may develop fatigue, arthralgias, and desquamation of the skin. Psychologic dependence has also been described. Even after the dose has been reduced to physiologic levels, HPA axis suppression (ie, secondary adrenal insufficiency) persists for an average of 9 to 10 months but may continue for as long as 1 to 2 years.

REFERENCES

General

Arlt W, Stewart PM. Adrenal corticosteroid biosynthesis, metabolism, and action. *Endocrinol Metab Clin North Am*. 2005;34:293. [PMID: 15850843]

Bottner A, Bornstein SR. Lessons learned from gene targeting and transgenesis for adrenal physiology and disease. *Rev Endocr Metab Disord*. 2001;2:275. [PMID: 11705133]

Finco I, LaPensee CR, Krill KT, Hammer GD. Hedgehog signaling and steroidogenesis. *Annu Rev Physiol*. 2015;77:105. [PMID: 25668018]

Gallo-Payet N, Battista MC. Steroidogenesis-adrenal cell signal transduction. *Compr Physiol*. 2014;4(3):889. [PMID: 24944026]

Greaves RF, Jevalikar G, Hewitt JK, Zacharin MR. A guide to understanding the steroid pathway: new insights and diagnostic implications. *Clin Biochem*. 2014;47(15):5. [PMID: 25086367]

Hammer GD, Parker KL, Schimmer BP. Minireview: transcriptional regulation of adrenocortical development. *Endocrinology*. 2005;146:1018. [PMID: 15604207]

Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. *Endocr Rev*. 2004;25:947. [PMID: 15131259]

Seckl JR. 11β-hydroxysteroid dehydrogenases: changing glucocorticoid action. *Curr Opin Pharm*. 2004;4:597. [PMID: 15525550]

Simard J, Ricketts ML, Gingras S, Soucy P, Feltus FA, Melner MH. Molecular biology of the 3β-hydroxysteroid dehydrogenase/Δ⁵-Δ⁴ isomerase gene family. *Endocr Rev*. 2005;26:525. [PMID: 15632317]

Spiga F, Lightman SL. Dynamics of adrenal glucocorticoid steroidogenesis in health and disease. *Mol Cell Endocrinol*. 2015;408:227. [PMID: 25662280]

Turcu AF, Auchus RJ. Adrenal steroidogenesis and congenital adrenal hyperplasia. *Endocrinol Metab Clin North Am*. 2015;44:275. [PMID: 26038201]

Xing Y, Lerario AM, Rainey W, Hammer GD. Development of adrenal cortex zonation. *Endocrinol Metab Clin North Am*. 2015;44:243. [PMID: 26038200]

Biologic Effects and Glucocorticoid Therapy

Chrousos GP. The glucocorticoid receptor gene, longevity, and the complex disorders of Western societies. *Am J Med*. 2004;117:204. [PMID: 15300973]

Czock D, Keller F, Rasche FM, Häussler U. Pharmacokinetics and pharmacodynamics of systemically administered glucocorticoids. *Clin Pharmacokinet*. 2005;44:61. [PMID: 15634032]

Limbourg FP, Liao JK. Nontranscriptional actions of the glucocorticoid receptor. *J Mol Med*. 2003;81:168. [PMID: 12682725]

Reichardt HM, Tronchea F, Bergera S, Kellendonka C, Schütza G. New insights into glucocorticoid and mineralocorticoid signaling: lessons from gene targeting. *Adv Pharmacol*. 2000;47:1. [PMID: 10582083]

Sizonenko PC. Effects of inhaled or nasal glucocorticosteroids on adrenal function and growth. *J Pediatr Endocrinol Metab*. 2002;15:5. [PMID: 11822580]

Tuckermann JP, Kleiman A, McPherson KG, Reichardt HM. Molecular mechanisms of glucocorticoids in the control of inflammation and lymphocyte apoptosis. *Crit Rev Clin Lab Sci*. 2005;42:71. [PMID: 15697171]

Vgontzas AN, Chrousos GP. Sleep, the hypothalamic-pituitary-adrenal axis, and cytokines: multiple interactions and disturbances in sleep disorders. *Endocrinol Metab Clin North Am*. 2002;31:15. [PMID: 12055986]

Adrenal Insufficiency

Annane D, Bellissant E, Bollaert PE, et al. Corticosteroids for severe sepsis and septic shock: a systematic review and meta-analysis. *BMJ*. 2004;329:480. [PMID: 15289273]

Arlt W, Allolio B. Adrenal insufficiency. *Lancet*. 2003;361:1881. [PMID: 12788587]

Boonen E, Bornstein SR, Van den Berghe G. New insights into the controversy of adrenal function during critical illness. *Lancet Diabetes Endocrinol*. 2015;3:805. [PMID: 26071883]

Bornstein SR. Predisposing factors for adrenal insufficiency. *N Engl J Med*. 2009;360:2328. [PMID: 19474430]

Bornstein SR, Allolio B, Arlt W, et al. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2016;10:364. [PMID: 26760044]

Broersen LH, Pereira AM, Jørgensen JO, Dekkers OM. Adrenal insufficiency in corticosteroids use: systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2015;100:2171. [PMID: 25844620]

Buttgereit F, Burmester GR, Lipworth BJ. Optimised glucocorticoid therapy: the sharpening of an old spear. *Lancet*. 2005;365:801. [PMID: 15733723]

Charmandari E, Nicolaides NC, Chrousos GP. Adrenal insufficiency. *Lancet*. 2014;383:2152. [PMID: 24503135]

Coursin DB, Wood KE. Corticosteroid supplementation for adrenal insufficiency. *JAMA*. 2002;287:236. [PMID: 11779267]

Eisenbarth GS, Gottlieb PA. Autoimmune polyendocrine syndromes. *N Engl J Med*. 2005;350:2068. [PMID: 15141045]

Fardet L, Fève B. Systemic glucocorticoid therapy: a review of its metabolic and cardiovascular adverse events. *Drugs*. 2014;74:1731. [PMID: 25204470]

Glowniak JV, Loriaux DL. A double-blind study of perioperative steroid requirements in secondary adrenal insufficiency. *Surgery*. 1997;121:123. [PMID: 9037222]

Guran T, Buonocore F, Saka N, et al. Rare causes of primary adrenal insufficiency: genetic and clinical characterization of a large nationwide cohort. *J Clin Endocrinol Metab*. 2016;101:284. [PMID: 26523528]

Hamrahian AH, Oseni TS, Arafah BM. Measurements of serum free cortisol in critically ill patients. *N Engl J Med*. 2004;350:1629. [PMID: 15084695]

Inder WJ, Hunt PJ. Glucocorticoid replacement in pituitary surgery: guidelines for perioperative assessment and management. *J Clin Endocrinol Metab*. 2002;87:2745. [PMID: 12050244]

Libè R, Barbetta L, Dall'Asta C, et al. Effects of dehydroepiandrosterone (DHEA) supplementation on hormonal, metabolic and behavioral status in patients with hypoadrenalism. *J Endocrinol Invest*. 2004;27:736. [PMID: 15636426]

Mallappa A, Debono M. Recent advances in hydrocortisone replacement treatment. *Endocr Dev*. 2016;30:42. [PMID: 26683495]

Olafsson AS, Sigurjonsdottir HA. Increasing prevalence of Addison disease: results from a nationwide study. *Endocr Pract*. 2016;22(1):30. [PMID: 26437215]

Ospina NS, Al Nofal A, Bancos I, et al. ACTH stimulation tests for the diagnosis of adrenal insufficiency: systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2016;101:427. [PMID: 26649617]

Richter B, Neises G, Clar C. Glucocorticoid withdrawal schemes in chronic medical disorders. A systematic review. *Endocrinol Metab Clin North Am*. 2002;31:751. [PMID: 12227130]

Sprung CL, Annane D, Keh D, et al. Hydrocortisone therapy for patients with septic shock. *N Engl J Med*. 2008;358:111. [PMID: 18184957]

Widmer IE, Puder JJ, König C, et al. Cortisol response in relation to the severity of stress and illness—the "CORESSI"-Study. *J Clin Endocrinol Metab*. 2005;90:4579. [PMID: 15886236]

Cushing Syndrome

Assie G, Libé R, Espiard S, et al. ARMC5 mutations in macronodular adrenal hyperplasia with Cushing's syndrome. *N Eng J Med*. 2013;369:2105. [PMID: 24283224]

Beuschlein F, Fassnacht M, Assié G, et al. Constitutive activation of PKA catalytic subunit in adrenal Cushing's syndrome. *N Eng J Med*. 2014;370:1019. [PMID: 24571724]

Di Dalmazi G, Vicennati V, Garelli S, et al. Cardiovascular events and mortality in patients with adrenal incidentalomas that are either non-secreting or associated with intermediate phenotype or subclinical Cushing's syndrome: a 15-year retrospective study. *Lancet Diabetes Endocrinol*. 2014;2:396. [PMID: 24795253]

Elamin MD, Murad MH, Mullan R, et al. Accuracy of diagnostic tests for Cushing's syndrome: a systematic review and metaanalyses. *J Clin Endocrinol Metab*. 2008;94:741. [PMID: 19088166]

Findling JW, Raff H. Screening and diagnosis of Cushing's syndrome. *Endocrinol Metab Clin North Am*. 2005;34:385. [PMID: 15850849]

Findling JW, Raff H, Aron DC. The low-dose dexamethasone suppression test: a reevaluation in patients with Cushing's syndrome. *J Clin Endocrinol Metab*. 2004;89:1222. [PMID: 15001614]

Hammer GD, Tyrrell JB, Lamborn KR, et al. Transsphenoidal microsurgery for Cushing's disease: initial outcome and long-term results. *J Clin Endocrinol Metab*. 2004;89:6348. [PMID: 15579802]

Iacobone M, Mantero F, Basso SM, Lumachi F, Favia G. Results and long-term follow-up after unilateral adrenalectomy for ACTH-independent hypercortisolism in a series of fifty patients. *J Endocrinol Invest*. 2005;28:327. [PMID: 15966505]

Isidori AM, Lenzi A. Ectopic ACTH syndrome. *Arq Bras Endocrinol Metab.* 2007;51:1217. [PMID: 18209859]

Nieman LK, Biller BM, Findling JW, et al. The diagnosis of Cushing's syndrome: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab.* 2008;93:1526. [PMID: 18334580]

Nieman LK, Biller BM, Findling JW, et al. Treatment of Cushing's Syndrome: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab.* 2015;100:2807. [PMID: 26222757]

Quddisi S, Browne P, Hirsch IB. Cushing's syndrome due to surreptitious glucocorticoid administration. *JAMA.* 1998;158:294. [PMID: 9472211]

Raff H, Carroll TB. Cushing's syndrome: from physiological principles to diagnosis and clinical care. *J Physiol.* 2015;593:493. [PMID: 25480800]

Savage MO, Lienhardt A, Lebrethon MC, et al. Cushing's disease in childhood: presentation, investigation, treatment and long-term outcome. *Horm Res.* 2001;55(suppl 1):24. [PMID: 11408758]

Stratakis CA, Kirschner LS, Carney JA. Clinical and molecular features of the Carney complex: diagnostic criteria and recommendations for patient evaluation. *J Clin Endocrinol Metab.* 2001;86:4041. [PMID: 11549623]

van Aken MO, Pereira AM, Biermasz NR, et al. Quality of life in patients after long-term biochemical cure of Cushing's disease. *J Clin Endocrinol Metab.* 2005;90:3279. [PMID: 15741267]

Yaneva M, Mosnier-Pudar H, Dugué MA, et al. Midnight salivary cortisol for the initial diagnosis of Cushing's syndrome of various causes. *J Clin Endocrinol Metab.* 2004;89:3345. [PMID: 15240613]

Hirsutism and Virilization

Azziz R. The evaluation and management of hirsutism. *Obstet Gynecol.* 2003;101:995. [PMID: 12738163]

Cabrera MS, Vogiatzi MG, New MI. Long term outcome in adult males with classic congenital adrenal hyperplasia. *J Clin Endocrinol Metab.* 2001;86:3070. [PMID: 11443169]

Kelestimur F. Hirsutism of adrenal origin in adolescents: consequences in adults. *J Pediatr Endocrinol Metab.* 2001;14:1309. [PMID: 11964028]

New MI. Inborn errors of adrenal steroidogenesis. *Mol Cell Endocrinol.* 2003;211:75. [PMID: 14656479]

Speiser PW, White PC. Congenital adrenal hyperplasia. *N Engl J Med.* 2003;349:776. [PMID: 12930931]

Turcu A, Smith JM, Auchus R, Rainey WE. Adrenal androgens and androgen precursors-definition, synthesis, regulation and physiologic actions. *Compr Physiol.* 2014;4:1369. [PMID: 25428847]

Incidentally Discovered Adrenal Masses and Adrenal Cancer

Aron DC, Howlett TA. Incidentalomas. *Endocrinol Metab Clin North Am.* 2000;29:205. [PMID: 10732272]

Caoili EM, Korobkin M, Francis IR, et al. Adrenal masses: characterization with combined unenhanced and delayed enhanced CT. *Radiology.* 2002;222:629. [PMID: 11867777]

Grumbach MM, Biller BMK, Braunstein GD, et al. Management of the clinically inapparent adrenal mass ("incidentaloma"). *Ann Intern Med.* 2003;138:424. [PMID: 12614096]

Nakamura Y, Yamazaki Y, Felizola SJ, et al. Adrenocortical carcinoma: review of the pathologic features, production of adrenal steroids, and molecular pathogenesis. *Endocrinol Metab Clin North Am.* 2015;44:399. [PMID: 26038208]

Sahdev A, Reznek RH. Imaging evaluation of the non-functioning indeterminate adrenal mass. *Trend Endocrinol Metab.* 2004;15:271. [PMID: 15358280]

Young WF. The incidentally discovered adrenal mass. *N Engl J Med.* 2007;356:601. [PMID: 17287480]

This page intentionally left blank

Endocrine Hypertension

William F. Young, Jr, MD, MSc

11β-HSD2	11β-Hydroxysteroid dehydrogenase type 2	**CT**	Computed tomography
ACE	Angiotensin-converting enzyme	**DOC**	Deoxycorticosterone
ACTH	Corticotropin	**FH**	Familial hyperaldosteronism
AME	Apparent mineralocorticoid excess	**GH**	Growth hormone
ANP	Atrial natriuretic peptide	**GRA**	Glucocorticoid-remediable aldosteronism
APA	Aldosterone-producing adenoma	**HU**	Hounsfield units
ARB	Angiotensin receptor blocker	**IHA**	Idiopathic hyperaldosteronism
AVS	Adrenal venous sampling	**IVC**	Inferior vena cava
BMAH	Bilateral macronodular adrenal hyperplasia	**PAC**	Plasma aldosterone concentration
CAH	Congenital adrenal hyperplasia	**PAH**	Primary adrenal hyperplasia
		PRA	Plasma renin activity

Hypertension affects one in four adults in the developed world. Although hypertension is *essential* or *idiopathic* in most cases, a cause can be detected in approximately 15% of the hypertensive population. The secondary causes of hypertension can be divided into renal (eg, renal vascular or parenchymal disease) and endocrine causes. There are at least 14 endocrine disorders in which hypertension may be the initial clinical presentation (Table 10–1). The diagnosis of endocrine hypertension presents the clinician an opportunity to provide a surgical cure or to achieve a marked response with targeted pharmacologic therapy. Pheochromocytoma and Cushing syndrome are reviewed in detail in Chapters 11 and 9, respectively. The renin-angiotensin-aldosterone system, the diagnostic and therapeutic approaches to mineralocorticoid hypertension (eg, primary aldosteronism), and less common forms of endocrine hypertension are reviewed in this chapter.

RENIN-ANGIOTENSIN-ALDOSTERONE SYSTEM

The components of the renin-angiotensin-aldosterone system are shown in Figure 10–1. Aldosterone is secreted from the zona glomerulosa under the primary control of angiotensin II, potassium, and corticotropin (ACTH). The secretion of aldosterone is restricted to the zona glomerulosa because of zonal-specific expression of aldosterone synthase (CYP11B2). Hypokalemia, atrial natriuretic peptide (ANP), dopamine, and heparin inhibit aldosterone secretion.

Renin and Angiotensin

Renin is an enzyme produced in the juxtaglomerular apparatus of the kidney, stored in granules, and released in response to specific secretagogues. The first 43 amino acids of the 340 amino acid renin protein are a prosegment cleaved to produce the active enzyme. The release of renin into the circulation is the rate-limiting step in the activation of the renin-angiotensin-aldosterone system. Renal renin release is controlled by: juxtaglomerular cells acting as pressure transducers that sense stretch of the afferent arteriolar wall and thus renal perfusion pressure; the macula densa, a specialized group of convoluted distal tubular cells that function as chemoreceptors for monitoring the sodium and chloride loads present in the distal tubule; and, the sympathetic nervous system, which modifies the release of renin, particularly in response to upright posture; and humoral factors, including potassium, angiotensin II, and ANPs. Thus, renin release is maximized in conditions of low renal perfusion pressure or low tubular sodium content (eg, renal artery stenosis, hemorrhage, dehydration). Renin release is suppressed by elevated perfusion pressure at the kidney (eg, hypertension) and high sodium diets.

TABLE 10–1 Endocrine causes of hypertension.

Adrenal Dependent
Pheochromocytoma
Primary aldosteronism
Hyperdeoxycorticosteronism
Congenital adrenal hyperplasia
11β-Hydroxylase deficiency
17α-Hydroxylase deficiency
Deoxycorticosterone-producing tumor
Primary cortisol resistance
Cushing syndrome
Apparent Mineralocorticoid Excess (AME)/11β-Hydroxysteroid Dehydrogenase Deficiency
Genetic
Type 1 AME
Acquired
Licorice or carbenoxolone ingestion (type 1 AME)
Cushing syndrome (type 2 AME)
Thyroid Dependent
Hypothyroidism
Hyperthyroidism
Pituitary Dependent
Acromegaly
Cushing disease

FIGURE 10–2 Steps in the production of angiotensin peptides by the renin-angiotensin system (ACE, angiotensin-converting enzyme).

Renin release is increased directly by hypokalemia and decreased by hyperkalemia.

Angiotensinogen, an α_2-globulin synthesized in the liver, is the substrate for renin and is broken down into the angiotensin peptides. Angiotensinogen consists of 485 amino acids, 33 of which constitute a signal peptide that is cleaved prior to secretion. The action of renin on angiotensinogen produces angiotensin I. Angiotensin I is composed of the first 10 amino acid sequence following the signal peptide and does not have biologic activity. Angiotensin II, the main form of biologically active angiotensin, is formed by cleavage of the two carboxyl-terminal amino acids of angiotensin I by angiotensin-converting enzyme (ACE) (Figure 10–2). ACE is localized to cell membranes in the lung and intracellular granules in certain tissues that produce angiotensin II. Amino peptidase A removes the amino-terminal aspartic acid to produce the heptapeptide, angiotensin III. Angiotensin II and angiotensin III have equivalent efficacy in promoting aldosterone secretion and modifying renal blood flow. The half-life in the circulation of angiotensin II is short (<60 seconds). Elements of the renin-angiotensin-aldosterone system are present in the adrenal glands, the kidneys, the heart, and the brain. For example, the adrenal glomerulosa cells contain the proteins needed to produce and secrete angiotensin II. Other tissues contain one or more components of the renin-angiotensin system and require other cells or circulating components, or both, to generate angiotensin II.
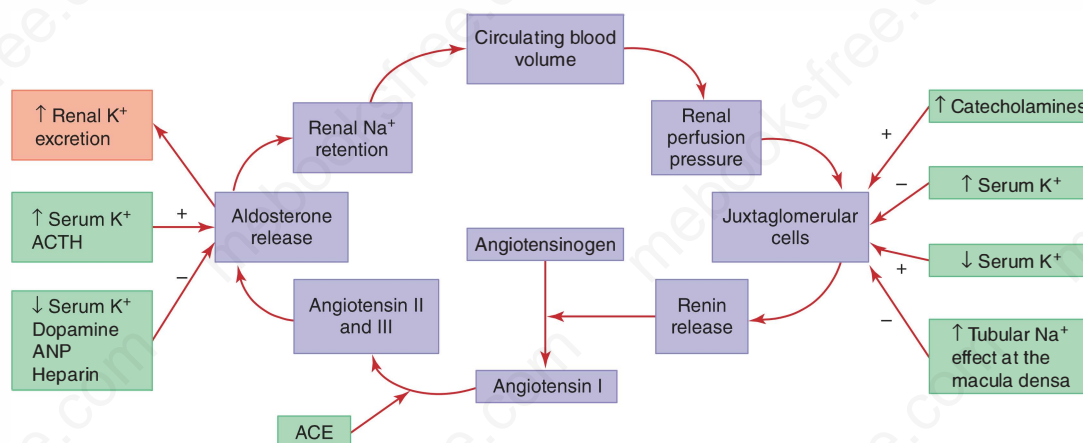
FIGURE 10–1 Renin-angiotensin-aldosterone and potassium-aldosterone feedback loops. Zona glomerulosa aldosterone production and secretion are determined by input from each loop (ACE, angiotensin-converting enzyme; ACTH, corticotropin; ANP, atrial natriuretic peptide; BP, blood pressure; K⁺, potassium; Na⁺, sodium).

Angiotensin II functions through the angiotensin receptor to maintain normal extracellular volume and blood pressure by: (a) increasing aldosterone secretion from the zona glomerulosa by increasing transcription of CYP11B2; (b) constriction of vascular smooth muscle, thereby increasing blood pressure and reducing renal blood flow; (c) enhancing the release of norepinephrine and epinephrine from the adrenal medulla; (d) enhancement of the activity of the sympathetic nervous system by increasing central sympathetic outflow, thereby increasing norepinephrine discharge from sympathetic nerve terminals; and, (e) promotion of the release of vasopressin.

Aldosterone

Aldosterone is produced in the zona glomerulosa of the adrenal cortex. Approximately 50% to 70% of aldosterone circulates bound to either albumin or corticosteroid-binding globulin; 30% to 50% of total plasma aldosterone is free. Aldosterone is rapidly inactivated to tetrahydroaldosterone in the liver and has a half-life of 15 to 20 minutes. Aldosterone regulates extracellular volume and potassium homeostasis by binding to renal cortical collecting duct principal epithelial cell mineralocorticoid receptors (Figure 10–3). The mineralocorticoid receptor—a member of the nuclear receptor family and also found in the heart, colon, and hippocampus—is localized to the cytoplasm prior to activation, undergoes a conformation change on binding to aldosterone, and translocates into the nucleus where it functions as a transcription factor. The aldosterone-regulated serum- and glucocorticoid-inducible kinase appears to be a key intermediary (see Figure 10–3). Aldosterone increases expression of this kinase which phosphorylates and inactivates neural-precursor-cell-expressed, developmentally down regulated (Nedd) 4-2, a ubiquitin ligase which is responsible for degrading the epithelial sodium channel. This, in turn, leads to an increased number of open sodium channels in the luminal membrane of the principal cells in the cortical collecting tubule, resulting in increased sodium ion reabsorption. The sodium loss increases luminal electronegativity, which augments tubular secretion of potassium by the renal tubular cells and hydrogen ion by the renal interstitial cells. Another mediator of the mineralocorticoid receptor transcriptional response is the activation is the sodium-potassium ATPase at the basolateral membrane, which drives the
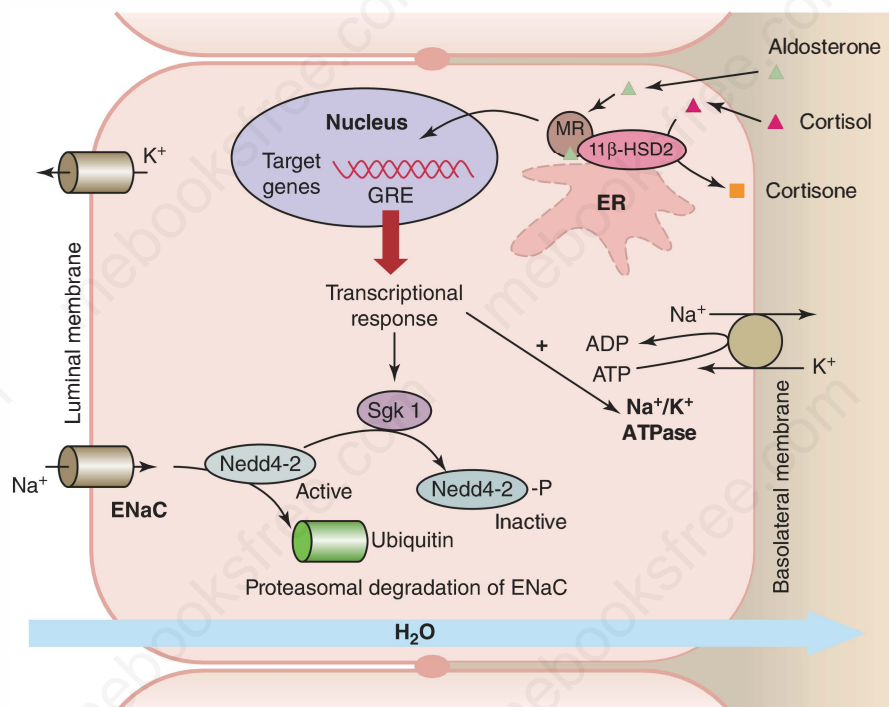
FIGURE 10–3 Aldosterone regulates extracellular volume and potassium homeostasis by binding to the renal cortical collecting duct principal epithelial cell mineralocorticoid receptor (MR). The activated MR translocates into the nucleus where it binds to the glucocorticoid response element (GRE) and functions as a transcription factor. Aldosterone increases expression of serum- and glucocorticoid-inducible kinase (Sgk1), which phosphorylates and inactivates neural-precursor-cell-expressed, developmentally downregulated gene (Nedd) 4-2, a ubiquitin ligase which is responsible for degrading the epithelial sodium channel (ENaC). Another mediator of mineralocorticoid receptor transcriptional response is the activation is the sodium-potassium ATPase (Na^+/K^+ ATPase) at the basolateral membrane, which drives the uptake of potassium and export of sodium. Although glucocorticoids and mineralocorticoids bind equally to the mineralocorticoid receptor, specificity of action is due to the glucocorticoid-degrading enzyme, 11β-hydroxysteroid dehydrogenase type 2 (11β-HSD2), which prevents glucocorticoids from interacting with the receptor. (Adapted with permission from Odermatt A, Atanasov AG. Mineralocorticoid receptors: emerging complexity and functional diversity. *Steroids.* 2009 Feb;74(2):163-171.)

uptake of potassium and export of sodium (see Figure 10–3). Although glucocorticoids and mineralocorticoids bind equally to the mineralocorticoid receptor, specificity of action is due to the glucocorticoid-degrading enzyme, 11β-hydroxysteroid dehydrogenase, which is strongly expressed in the kidney and prevents glucocorticoids from interacting with the receptor.

Aldosterone has nonclassic effects that, although probably genomic and therefore mediated by activation of the cytosolic mineralocorticoid receptor, do not include modification of sodium-potassium balance. Aldosterone-mediated actions include: the expression of several collagen genes; activation of genes controlling tissue growth factors, such as transforming growth factor β and plasminogen activator inhibitor type 1; and, increased expression of genes mediating inflammation. The resultant actions lead to microangiopathy, necrosis (acutely), and fibrosis in various tissues such as the heart, the vasculature, and the kidney. Increased levels of aldosterone are not necessary to cause this damage; an imbalance between the volume or sodium balance state and the level of aldosterone appear to be the critical factors. Spironolactone and eplerenone are mineralocorticoid receptor antagonists. Mineralocorticoid receptor blockade has proven to be clinically important in patients with cardiovascular disease. For example, when spironolactone was added to the treatment program for patients with New York Heart Association class IV heart failure or class III heart failure, it resulted in a significant 30% reduction in overall mortality due to reductions in death from heart failure and sudden death. When eplerenone was added to the treatment program for patients who had a myocardial infarction 3 to 14 days previously and had a left ventricular ejection fraction of less than or equal to 40%, it resulted in a significantly lower rate of cardiovascular mortality and sudden cardiac death. The effect of mineralocorticoid receptor antagonists on survival in patients with primary aldosteronism has not yet been studied.

The action of angiotensin II on aldosterone synthesis and secretion involves a feedback loop that also includes extracellular fluid volume (see Figure 10–1). A decrease in circulating blood volume results in decreased renal perfusion pressure that is detected by the renal juxtaglomerular cells. Activation of the juxtaglomerular cells increases renin release, which catalyzes the conversion of angiotensinogen to angiotensin I. ACE in the pulmonary and renal endothelium catalyzes the conversion of angiotensin I to angiotensin II and III, which act on the adrenal zona glomerulosa angiotensin receptor to stimulate aldosterone release. Aldosterone acts at the renal mineralocorticoid receptors to stimulate sodium and water retention to preserve the circulating blood volume. Renin release can also be triggered by catecholamines, hypokalemia, and a decrease in sodium chloride absorption in the macula densa cells. Aldosterone secretion can be directly stimulated by ACTH and hyperkalemia. Thus, sodium restriction activates and sodium overload suppresses the renin-angiotensin-aldosterone axis. Mineralocorticoid escape refers to the counterregulatory mechanisms that occur after 3 to 5 days of excessive mineralocorticoid administration. Several mechanisms contribute to this escape, including renal hemodynamic factors and increased release of ANP.

Excess aldosterone secretion causes hypertension through two main mechanisms: (1) mineralocorticoid-induced expansion of plasma and extracellular fluid volume; and (2) increase in total peripheral vascular resistance.

PRIMARY ALDOSTERONISM

Hypertension, suppressed plasma renin activity (PRA), and increased aldosterone secretion characterize the syndrome of primary aldosteronism. Aldosterone-producing adenoma (APA) and bilateral idiopathic hyperaldosteronism (IHA) are the two most common subtypes of primary aldosteronism (Table 10–2). Somatic mutations account for about half of APAs and include mutations in genes encoding components of: the Kir 3.4 (GIRK4) potassium channel (KCNJ5); the sodium/potassium and calcium ATPases (ATP1A1 and ATP2B3); and a voltage-dependent C-type calcium channel (CACNA1D). A much less common form, unilateral hyperplasia, is caused by micronodular or macronodular

TABLE 10–2 Adrenocortical causes of hypertension.

Low Renin and High Aldosterone
PRIMARY ALDOSTERONISM
Aldosterone-producing adenoma (APA) ≈35% of cases
Bilateral idiopathic hyperplasia (IHA) ≈60% of cases
Unilateral (primary) adrenal hyperplasia ≈2% of cases
Aldosterone-producing adrenocortical carcinoma <1% of cases
Familial hyperaldosteronism (FH)
    Glucocorticoid-remediable aldosteronism (FH type I) <1% of cases
    FH type II (APA or IHA) <2% of cases
    FH type III (associated with the germline mutation in the KCNJ5 potassium channel): <1% of cases
Ectopic aldosterone-producing adenoma or carcinoma <0.1% of cases

Low Renin and Low Aldosterone
HYPERDEOXYCORTICOSTERONISM
    Congenital adrenal hyperplasia
        11β-Hydroxylase deficiency
        17α-Hydroxylase deficiency
    Deoxycorticosterone-producing tumor
    Primary cortisol resistance

APPARENT MINERALOCORTICOID EXCESS (AME)/11β-HYDROXYSTEROID DEHYDROGENASE DEFICIENCY
Genetic
    Type 1 AME
    Type 2 AME
Acquired
    Licorice or carbenoxolone ingestion (type 1 AME)
    Cushing syndrome (type 2 AME)

CUSHING SYNDROME
Exogenous glucocorticoid administration—most common cause
Endogenous
    ACTH-dependent ≈85% of cases
        Pituitary
        Ectopic
    ACTH-independent ≈15% of cases
        Unilateral adrenal disease
        Bilateral adrenal disease
            Bilateral macronodular adrenal hyperplasia (rare)
            Primary pigmented nodular adrenal disease (rare)

Abbreviations: ACTH, corticotropin; AME, apparent mineralocorticoid excess; APA, aldosterone-producing adenoma; FH, familial hyperaldosteronism; IHA, idiopathic hyperaldosteronism.

hyperplasia of the zona glomerulosa of predominantly one adrenal gland. Unilateral hyperplasia is referred to as primary adrenal hyperplasia (PAH). Familial hyperaldosteronism (FH) is also rare and three types have been described: FH types 1, 2, and 3.

FH type 1, or glucocorticoid-remediable aldosteronism (GRA), is caused by a chimeric gene duplication that results from unequal crossing over between the promoter sequence of *CYP11B1* gene (encoding 11β-hydroxylase) and the coding sequence of CYP11B2 (encoding aldosterone synthase). This chimeric gene contains the 3' ACTH-responsive portion of the promoter from the 11β-hydroxylase gene fused to the 5' coding sequence of the aldosterone synthase gene. The result is ectopic expression of aldosterone synthase activity in the cortisol-producing zona fasciculata. Thus, mineralocorticoid production is regulated by ACTH instead of the normal secretagogue, angiotensin II. Aldosterone secretion can be suppressed by glucocorticoid therapy.

FH type 2 (FH-2) refers to the familial occurrence of APA or IHA or both. FH-2 is autosomal dominant and may be monogenic. The hyperaldosteronism in FH-2 does not suppress with dexamethasone, and GRA mutation testing is negative. FH-2 is more common than FH-1, but it still accounts for fewer than 3% of all patients with primary aldosteronism. The molecular basis for FH-2 is unclear, although a linkage analysis study showed an association with chromosomal region 7p22.

FH-3 was first described in a single family in 2008. This initial report included a father and two daughters who all presented with refractory hypertension before age 7 years and all three were treated with bilateral adrenalectomy. The adrenal glands showed massive hyperplasia. Three years later the causative germline mutation in this family was discovered: a point mutation in and near the selectivity filter of the potassium channel *KCNJ5*. This *KCNJ5* mutation produces increased sodium conductance and cell depolarization, triggering calcium entry into glomerulosa cells, the signal for aldosterone production and cell proliferation. Other families with early onset hyperaldosteronism have also been identified to have germline point mutations in the *KCNJ5* gene.

Prevalence

In the past, clinicians would not consider the diagnosis of primary aldosteronism unless the patient presented with spontaneous hypokalemia, and then the diagnostic evaluation would require discontinuing antihypertensive medications for at least 2 weeks. The *spontaneous hypokalemia/no antihypertensive drug* diagnostic approach resulted in predicted prevalence rates of less than 0.5% of hypertensive patients. However, it is now recognized that most patients with primary aldosteronism are not hypokalemic and that case-detection testing can be completed with a simple blood test (plasma aldosterone concentration [PAC]-to-plasma renin activity [PRA] ratio) while the patient is taking most antihypertensive drugs. Using the PAC-PRA ratio as a case-detection test, followed by aldosterone suppression confirmatory testing, has resulted in much higher prevalence estimates (5%-10% of all patients with hypertension) for primary aldosteronism.

Clinical Presentation

The diagnosis of primary aldosteronism is usually made in patients who are in the third to sixth decade of life. Few symptoms are specific to the syndrome. Patients with marked hypokalemia may have muscle weakness and cramping, headaches, palpitations, polydipsia, polyuria, nocturia, or a combination of these. Periodic paralysis is a very rare presentation in Caucasians, but it is not an infrequent presentation in patients of Asian descent. The polyuria and nocturia are a result of a hypokalemia-induced renal concentrating defect and the presentation is frequently mistaken for prostatism in men. There are no specific physical findings. Edema is not a common finding because of mineralocorticoid escape. The degree of hypertension is usually moderate to severe and may be resistant to usual pharmacologic treatments. Although not common, primary aldosteronism may present with hypertensive urgencies. Patients with APA tend to have higher blood pressures than those with IHA. Hypokalemia is frequently absent; thus, all patients with hypertension are candidates for this disorder. In other patients, the hypokalemia only becomes evident with addition of a potassium-wasting diuretic (eg, hydrochlorothiazide, furosemide). Aldosterone excess also leads to a mild metabolic alkalosis because of increased urinary hydrogen excretion mediated both by hypokalemia and by the direct stimulatory effect of aldosterone on distal renal tubule acidification. Because of a reset osmostat, the serum sodium concentration tends to be high-normal or slightly above the upper limit of normal—this clinical finding is very useful when initially assessing the potential for primary aldosteronism.

Several studies have shown that patients with primary aldosteronism may be at higher risk than other patients with hypertension for target organ damage of the heart and kidney. When matched for age, blood pressure, and duration of hypertension, patients with primary aldosteronism have greater left ventricular mass by echocardiographic measurements than patients with other types of hypertension (eg, pheochromocytoma, Cushing syndrome, or essential hypertension). In patients with APA, the left ventricular wall thickness and mass decreases markedly 1 year after adrenalectomy. Patients presenting with either APA or IHA have a significantly higher rate of cardiovascular events (eg, stroke, atrial fibrillation, and myocardial infarction) than matched patients with essential hypertension who have similar degrees of hypertension duration and control.

Diagnosis

The diagnostic approach to primary aldosteronism can be considered in three parts: (A) case-detection tests, (B) confirmatory tests, and (C) subtype evaluation tests.

A. Case-detection tests Spontaneous hypokalemia is uncommon in patients with uncomplicated hypertension and, when present, strongly suggests associated mineralocorticoid excess. However, most patients with primary aldosteronism have baseline serum levels of potassium in the normal range. Therefore, hypokalemia is not and should not be the only criterion

Consider testing for primary aldosteronism in the following settings:
• Hypertension and hypokalemia
• Resistant hypertension (three drugs and poor BP control)
• Adrenal incidentaloma and hypertension
• Onset of hypertension at a young age (<30 y)
• Severe hypertension (≥150 mm Hg systolic or ≥100 mm Hg diastolic)
• Whenever considering secondary hypertension

Case-detection testing:
Morning blood sample in seated ambulant patient
• Plasma aldosterone concentration (PAC)
• Plasma renin activity (PRA) or plasma renin concentration (PRC)

↑ PAC (≥15 ng/dL; ≥416 pmol/L)
↓ PRA (<1.0 ng/mL/h)
and
PAC-PRA ratio ≥20 ng/dL per ng/mL/h (≥555 pmol/L per ng/mL/h)

Confirmatory testing

FIGURE 10–4 When to consider testing for primary aldosteronism and use of the plasma aldosterone concentration to plasma renin activity ratio as a case-finding tool (BP, blood pressure; PAC, plasma aldosterone concentration; PRA, plasma renin activity; PRC, plasma renin concentration).

used to determine whom to test for primary aldosteronism. Patients with hypertension and hypokalemia (regardless of presumed cause), treatment-resistant hypertension (three antihypertensive drugs and poor control), severe hypertension (≥150 mm Hg systolic or ≥100 mm Hg diastolic), hypertension and an incidental adrenal mass, and onset of hypertension at a young age (eg, <30 years) should undergo case-detection testing for primary aldosteronism (Figure 10–4). In addition, primary aldosteronism should be tested for when considering a secondary hypertension evaluation (eg, when testing for renovascular disease or pheochromocytoma).

In patients with suspected primary aldosteronism, case detection can be accomplished by measuring a morning (preferably between 8 AM and 10 AM) ambulatory paired random PAC and PRA (see Figure 10–4). This test may be performed while the patient is taking most antihypertensive medications and without posture stimulation. Hypokalemia reduces the secretion of aldosterone, and it is optimal in patients with hypokalemia to restore the serum level of potassium to normal before performing diagnostic studies. It may be difficult to interpret data obtained from patients treated with a mineralocorticoid receptor antagonist (spironolactone and eplerenone). These drugs prevent aldosterone from activating the receptor, resulting sequentially in sodium loss, a decrease in plasma volume, and an elevation in PRA, which will reduce the utility of the PAC/PRA ratio. For this reason, spironolactone and eplerenone should not be initiated until the evaluation is completed, and the final decisions about treatment are made. However, there are rare exceptions to this rule. For example, if the patient is hypokalemic despite treatment

with spironolactone or eplerenone, then the mineralocorticoid receptors are not fully blocked, and PRA should be suppressed in such a patient with primary aldosteronism. In this unique circumstance, the evaluation for primary aldosteronism can proceed, despite treatment with mineralocorticoid receptor antagonists. However, in most patients already receiving spironolactone, therapy should be discontinued for at least 6 weeks. Other potassium-sparing diuretics, such as amiloride and triamterene, usually do not interfere with testing unless the patient is on high doses.

ACE-inhibitors and angiotensin receptor blockers (ARBs) have the potential to falsely elevate PRA. Therefore, in a patient treated with an ACE inhibitor or ARB, the finding of a detectable PRA level or a low PAC-PRA ratio does not exclude the diagnosis of primary aldosteronism. However, a very useful clinical point is that when a PRA level is undetectably low in a patient taking an ACE inhibitor or ARB, primary aldosteronism is likely. A second important clinical point is that PRA is suppressed (<1.0 ng/mL/h) in almost all patients with primary aldosteronism, regardless of concurrent medications.

The PAC-PRA ratio is based on the concept of paired hormone measurements. For example, in a hypertensive hypokalemic patient: (a) secondary hyperaldosteronism should be considered when both PRA and PAC are increased and the PAC-PRA ratio is less than 10 (eg, renovascular disease); (b) an alternate source of mineralocorticoid receptor agonism should be considered when both PRA and PAC are suppressed (eg, hypercortisolism); and, (c) primary aldosteronism should be suspected when PRA is suppressed (<1.0 ng/mL/h) and PAC is increased (Figure 10–5). Although there is some uncertainty about test characteristics and lack of standardization, the PAC-PRA ratio is widely accepted as the case-detection test of choice for primary aldosteronism. It is important to understand that the lower limit of detection varies among different PRA assays and can have a dramatic effect on the PAC-PRA ratio. As an example, if the lower limit of detection for PRA is 0.6 ng/mL/h and the PAC is 18 ng/dL, then the PAC-PRA ratio would be 30; however, if the lower limit of detection for PRA is 0.1 ng/mL/h and the PAC is 18 ng/dL, then the PAC-PRA ratio would be 180. Thus, the cutoff for a high PAC-PRA ratio is laboratory dependent and, more specifically, PRA assay dependent. At Mayo Clinic, a PAC (in ng/dL)-PRA (in ng/mL/h) ratio of 20 or more and PAC of at least 15 ng/dL are found in more than 90% of patients with surgically confirmed APA. In patients without primary aldosteronism, most of the variation occurs within the normal range. The sensitivity and specificity of the PAC-PRA ratio in the diagnosis of primary aldosteronism are approximately 80% and 75%, respectively. A high PAC-PRA ratio with a PAC of at least 15 ng/dL is a positive case-detection test result, a finding that warrants further testing. Other initial case-detection strategies include measurement of isolated plasma renin activity or 24-hour urinary aldosterone excretion.

B. Confirmatory tests   An increased PAC-PRA ratio is not diagnostic by itself, and primary aldosteronism must be confirmed by demonstrating lack of normal suppressibility of aldosterone secretion. The list of drugs and hormones capable of affecting the

Hypertension and hypokalemia

Plasma aldosterone concentration (PAC)
Plasma renin activity (PRA)

↑ PAC
↑ PRA
PAC-PRA ratio <10
(277 in SI units)

↑ PAC
↓ PRA
PAC-PRA ratio ≥20
(≥555 in SI units)
and
PAC ≥15 ng/dL (≥416 pmol/L)

↓ PAC
↓ PRA

Renovascular hypertension
Diuretic use
Renin-secreting tumor
Malignant-phase hypertension
Coarctation of the aorta

Primary
aldosteronism

Congenital adrenal hyperplasia
Exogenous mineralocorticoid
DOC-producing tumor
Cushing syndrome
11 β-OHSD deficiency
Altered aldosterone metabolism
Liddle syndrome

FIGURE 10–5 Use of the plasma aldosterone concentration-to-plasma renin activity ratio to differentiate among different causes of hypertension and hypokalemia (DOC, deoxycorticosterone; OHSD, hydroxysteroid dehydrogenase; PAC, plasma aldosterone concentration; PRA, plasma renin activity; SI units, International System of Units). (Adapted with permission from Young WF Jr, Hogan MJ. Renin-independent hypermineralocorticoidism. *Trends Endocrinol Metab.* 1994 Apr;5(3):97-106.)

renin-angiotensin-aldosterone axis is extensive, and frequently, a medication-contaminated evaluation is unavoidable. Calcium channel blockers and α_1-adrenergic receptor blockers do not affect the diagnostic accuracy in most cases. It is impossible to interpret data obtained from patients receiving treatment with mineralocorticoid receptor antagonists (eg, spironolactone, eplerenone) when PRA is not suppressed. Therefore, treatment with a mineralocorticoid receptor antagonist should not be initiated until the evaluation has been completed and the final decisions about treatment have been made. Aldosterone suppression testing can be performed with orally administered sodium chloride and measurement of urinary aldosterone or with intravenous sodium chloride loading and measurement of PAC.

a. **Oral sodium loading test**—After hypertension and hypokalemia are controlled, patients should receive a high sodium diet (supplemented with sodium chloride tablets if needed) for 3 days, with a goal sodium intake of 5000 mg (equivalent to 12.8 g sodium chloride or 218 mEq of sodium). The risk of increasing dietary sodium in patients with severe hypertension must be assessed in each case. Because the high sodium diet can increase kaliuresis and hypokalemia, vigorous replacement of potassium chloride may be needed, and the serum level of potassium should be monitored daily. On the third day of the high sodium diet, a 24-hour urine specimen is collected for measurement of aldosterone, sodium, and creatinine. To document adequate sodium repletion, the 24-hour urinary sodium excretion should exceed 200 mEq. Urinary aldosterone excretion more than 12 µg/24 h is consistent with autonomous aldosterone secretion. The sensitivity and specificity of the oral sodium loading test are 96% and 93%, respectively.

b. **Intravenous saline infusion test**—The intravenous saline infusion test has also been used for the confirmation of primary aldosteronism. Normal subjects show suppression of PAC after volume expansion with isotonic saline; subjects with primary aldosteronism do not show this suppression. The risks associated with rapid intravenous volume expansion should be assessed in each case. The test is done after an overnight fast. Two liters of 0.9% sodium chloride solution are infused intravenously with an infusion pump over 4 hours into the recumbent patient. Blood pressure and heart rate are monitored during the infusion. At the completion of the infusion, blood is drawn for measurement of PAC. PAC levels in normal subjects decrease to less than 5 ng/dL; most patients with primary aldosteronism do not suppress to less than 10 ng/dL; post-saline infusion PAC values between 5 and 10 ng/dL are indeterminate and can be seen in patients with IHA.

C. Subtype studies   Following case detection and confirmatory testing, the third management step guides the therapeutic approach by distinguishing APA and PAH from IHA and GRA. Unilateral adrenalectomy in patients with APA or PAH results in normalization of hypokalemia in all; hypertension is improved in all and is cured in approximately 30% to 60% of them. In IHA and GRA, unilateral or bilateral adrenalectomy seldom corrects the hypertension. IHA and GRA should be treated medically. APA is found in approximately 35% of cases and bilateral IHA in approximately 60% of cases (see Table 10–2). APAs are usually hypodense nodules (<2 cm in diameter) on CT and are golden yellow in color on cut section. IHA adrenal glands may be normal on CT or show nodular changes. Aldosterone-producing adrenal carcinomas are almost always greater than 4 cm in diameter and have a suspicious imaging phenotype on CT. Patients with aldosterone-producing adrenocortical carcinomas usually have severe aldosterone excess with serum potassium concentrations frequently less than 2.0 mEq/L.
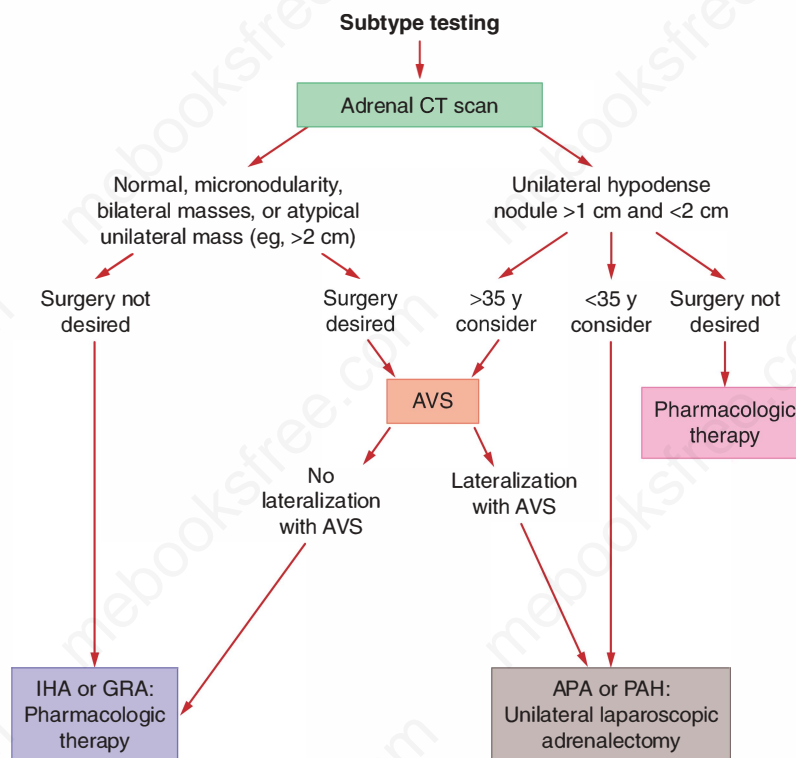
FIGURE 10–6 Subtype evaluation of primary aldosteronism. For patients who want to pursue a surgical treatment for their hypertension, adrenal venous sampling is frequently a key diagnostic step. See text for details (APA, aldosterone-producing adenoma; AVS, adrenal venous sampling; CT, computed tomography; IHA, idiopathic hyperaldosteronism; PAH, primary adrenal hyperplasia). (Adapted with permission from Young WF Jr, Hogan MJ. Renin-independent hypermineralocorticoidism. *Trends Endocrinol Metab*. 1994 Apr;5(3):97-106.)

a. **Adrenal CT**—Primary aldosteronism subtype evaluation may require one or more tests, the first of which is imaging the adrenal glands with CT (Figure 10–6). This imaging test is usually ordered as CT of the abdomen limited to the adrenal glands with 2-mm contiguous cuts. Although contrast enhancement is not necessary, contrast administration results in improved discrimination between normal adrenal cortex and the small lipid-rich cortical adenoma. When a solitary unilateral hypodense (Hounsfield units [HU] <10) macroadenoma (>1 cm) and normal contralateral adrenal morphology are found on CT in a young patient (adrenal incidentalomas are uncommon in patients <35 years) with primary aldosteronism, unilateral adrenalectomy is a reasonable therapeutic option (Figure 10–7). However, in most cases, CT may show normal-appearing adrenals, minimal unilateral adrenal limb thickening, unilateral microadenomas (<1 cm), bilateral macroadenomas, or a large (eg, >2 cm) unilateral macroadenomas that would be atypical for primary aldosteronism. In these cases, when the patient wants to pursue the surgical treatment option for primary aldosteronism, additional testing is required to determine the source of excess aldosterone secretion (Figure 10–8). Small APAs may be labeled incorrectly as IHA on the basis of CT findings of bilateral nodularity or normal-appearing adrenals. Also, apparent adrenal microadenomas may actually represent areas of hyperplasia, and unilateral adrenalectomy would be inappropriate. In addition, nonfunctioning unilateral adrenal macroadenomas are not uncommon, especially in individuals more than 35 years of age.

Unilateral PAH may be visible on CT or the PAH adrenal may appear normal on CT. In general, patients with APAs have more severe hypertension, more frequent hypokalemia, higher plasma (>30 ng/dL) and urinary (>30 µg/24 h) levels of aldosterone than those with IHA.

Adrenal CT is not accurate in distinguishing between APA and IHA. In one study of 203 patients with primary aldosteronism who were evaluated with both CT and adrenal venous sampling (AVS), CT was accurate in only 53% of patients; based on CT findings, 42 patients (22%) would have been incorrectly excluded as candidates for adrenalectomy and 48 (25%) might have had unnecessary or inappropriate surgery. Therefore, AVS is essential to direct appropriate therapy in patients older than 35 years of age with primary aldosteronism who have a high probability of APA and who seek a potential surgical cure. However, it is important to recognize that the surgical option is not mandatory in patients with APA—pharmacologic therapy with a mineralocorticoid receptor antagonist is the medication equivalent of adrenalectomy (discussed later).

b. **Adrenal venous sampling**—AVS is the criterion standard test to distinguish between unilateral and bilateral disease in patients with primary aldosteronism who want to pursue surgical management for their hypertension. AVS is an intricate procedure because the right adrenal vein is small and may be difficult to locate and cannulate—the success rate depends on the proficiency of the angiographer. The five keys to a successful AVS program are: (1) appropriate patient selection; (2) careful patient preparation; (3) focused
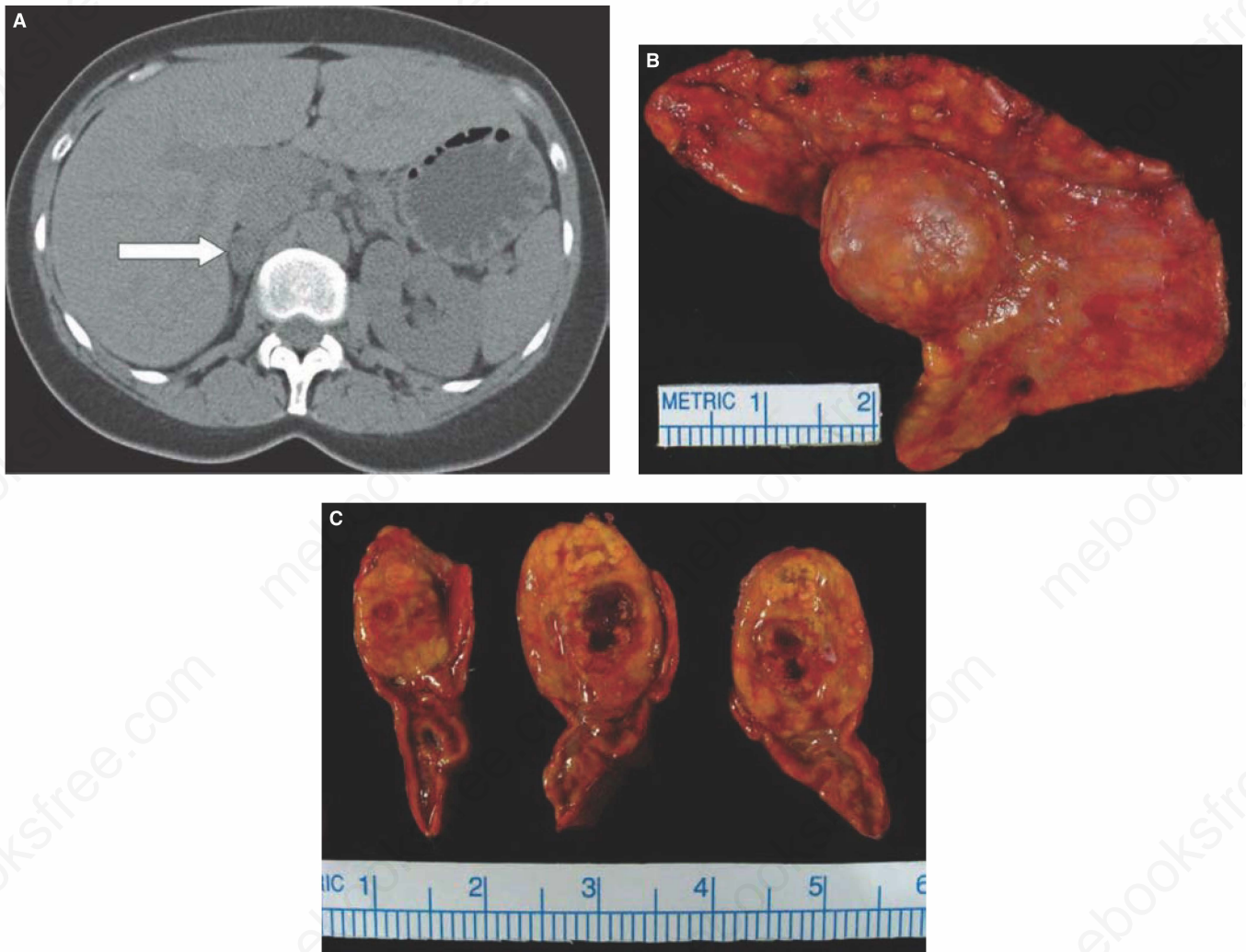
FIGURE 10–7 A 30-year-old woman had a 7-year history of hypertension and hypokalemia. Her blood pressure was not well controlled despite a four-drug program that included a calcium channel blocker, ACE-inhibitor, thiazide diuretic, and a β-adrenergic blocker. To correct her hypokalemia, she took 420 mEq of potassium per day. The case-detection test for primary aldosteronism was positive, with a plasma aldosterone concentration (PAC) of 34 ng/dL and low plasma renin activity (PRA) at less than 0.6 ng/mL/h (PAC-PRA ratio >56). The confirmatory test for primary aldosteronism was also positive, with 24-hour urinary excretion of aldosterone of 77 µg on a high-sodium diet (urinary sodium, 205 mEq/24 h). **A.** Adrenal CT axial image showing a 15-mm low-density mass (arrow) within the right adrenal gland. She underwent laparoscopic right adrenalectomy. **B.** Right adrenal gland (6.2 g, 6.1 cm × 3.7 cm × 1.4 cm) with a 1.8-cm cortical adenoma arising from the surface. **C.** Cut sections of the yellow adrenal cortical adenoma forming a 1.8 cm × 1.7 cm × 1.3 cm nodule. The postoperative plasma aldosterone concentration was less than 1.0 ng/dL. Hypokalemia was resolved. Four years after surgery, her blood pressure was normal without the aid of antihypertensive medications.

technical expertise; (4) defined protocol; and, (5) accurate data interpretation. A center-specific, written protocol is mandatory. The protocol should be developed by an interested group of endocrinologists, hypertension specialists, internists, radiologists, and laboratory personnel. Safeguards should be in place to prevent mislabeling of the blood tubes in the radiology suite and to prevent sample mix-up in the laboratory. At Mayo Clinic, we use continuous cosyntropin infusion during AVS (50 µg/h started 30 minutes before sampling and continued throughout the procedure) for the following reasons: (a) to minimize stress-induced fluctuations in aldosterone secretion during nonsimultaneous AVS; (b) to maximize the gradient in cortisol from adrenal vein to inferior vena cava (IVC) and thus confirm successful sampling of the adrenal veins; and, (c) to maximize the secretion of aldosterone from an APA. The adrenal veins are catheterized through the percutaneous femoral vein approach, and the position of the catheter tip is verified by gentle injection of a small amount of nonionic contrast medium and radiographic documentation (see Figure 10–8). Blood is obtained from both adrenal veins and the IVC below the renal veins and assayed for aldosterone and cortisol concentrations. To be sure there is no cross-contamination, the IVC sample should be obtained from the external iliac vein. The venous sample from the left side typically is obtained from the common phrenic
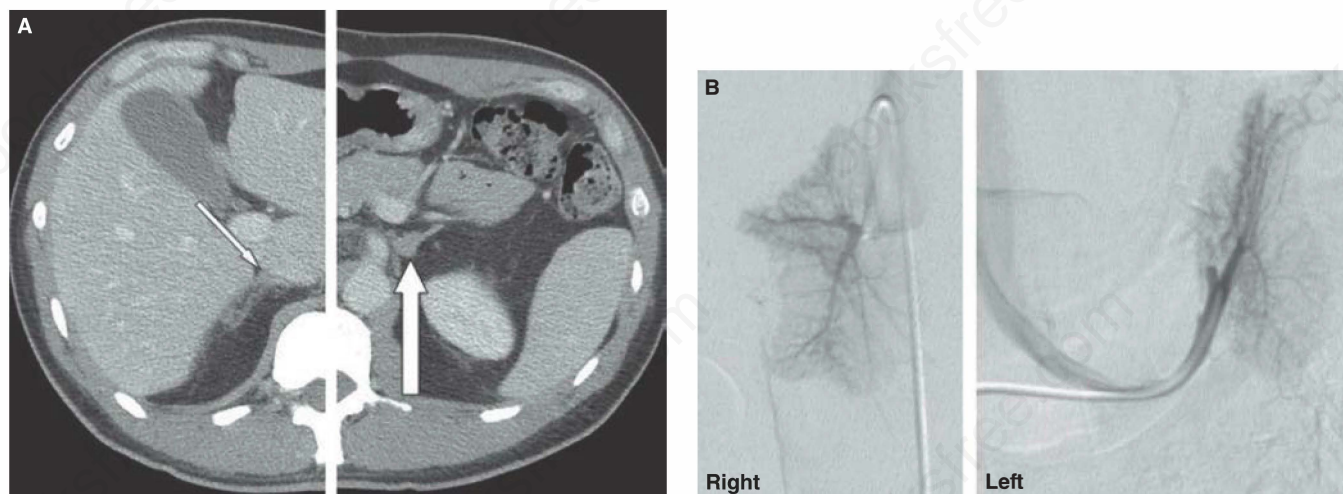
Bilateral adrenal venous sampling

Vein	Aldosterone (A), ng/dL	Cortisol (C), μg/dL	A/C ratio	Aldosterone ratio*
R adrenal vein	8450	1669	5.1	8.5
L adrenal vein	923	1599	0.6	
Inferior vena cava	13	22	0.6	

*R adrenal vein A/C ratio divided by L adrenal vein A/C ratio.

C

FIGURE 10–8 A 48-year-old man had a 7-year history of hypertension that was not optimally controlled on four antihypertensive drugs (β-adrenergic blocker, peripheral α$_1$-antagonist, angiotensin receptor blocker, and a thiazide diuretic). He was not hypokalemic. Resistant hypertension prompted case-detection testing for primary aldosteronism with a plasma aldosterone concentration (PAC) of 15 ng/dL and low plasma renin activity (PRA) at less than 0.6 ng/mL/h (PAC-PRA ratio >25). The confirmatory test for primary aldosteronism was also positive, with 24-hour urinary excretion of aldosterone of 16 μg on a high-sodium diet (urinary sodium, 356 mEq/24 h). **A.** Adrenal CT split section axial images show a 12-mm thickening (large arrow) in the inferior aspect of the left adrenal gland and a tiny nodule (small arrow) in the right adrenal gland. The patient wanted to pursue a surgical approach to the resolution of or improvement in hypertension. **B.** Adrenal venous sampling images showing the catheter in the right and left adrenal veins. The delicate venous architecture is demonstrated. **C.** Adrenal venous sampling lateralized aldosterone secretion to the right adrenal gland, and a 3-mm yellow cortical adenoma was found at laparoscopic right adrenalectomy. The postoperative plasma aldosterone concentration was less than 1.0 ng/dL. One year after surgery, his blood pressure was in the normal range with the aid of one antihypertensive medication. (Reproduced with permission from Young WF. Endocrine hypertension: then and now. *Endocr Pract*. 2010 Sep-Oct;16(5):888-902.)

vein immediately adjacent to the entrance of the adrenal vein. The cortisol concentrations from the adrenal veins and IVC are used to confirm successful catheterization; the adrenal vein-IVC cortisol ratio is typically more than 10:1.

Dividing the right and left adrenal vein PACs by their respective cortisol concentrations corrects for the dilutional effect of the inferior phrenic vein flow into the left adrenal vein; these are termed **cortisol-corrected ratios** (see Figure 10–8). In patients with APA, the mean cortisol-corrected aldosterone ratio (APA-side PAC/cortisol:normal adrenal PAC/cortisol) is 18:1. A cutoff of the cortisol-corrected aldosterone ratio from high side to low side more than 4:1 is used to indicate unilateral aldosterone excess (see Figure 10–8). In patients with IHA, the mean cortisol-corrected aldosterone ratio is 1.8:1 (high side:low side); a ratio less than 3:1 is suggestive of bilateral aldosterone hypersecretion. Therefore, most patients with a unilateral source of aldosterone will have cortisol-corrected aldosterone

lateralization ratios greater than 4.0; ratios greater than 3.0 but less than 4.0 represent a zone of overlap. Ratios no more than 3.0 are consistent with bilateral aldosterone secretion. The test characteristics of AVS for detecting unilateral aldosterone hypersecretion (APA or PAH) have sensitivity of 95% and specificity of 100%. However, since all patients who undergo AVS are not sent to surgery, the true diagnostic sensitivity of AVS is unknown. At centers with experience with AVS, the complication rate is 2.5% or less. Complications can include symptomatic groin hematoma, adrenal hemorrhage, and dissection of an adrenal vein.

Some centers and clinical practice guidelines recommend that AVS should be performed in all patients who have the diagnosis of primary aldosteronism. However, a more practical approach is to consider the use of AVS based on patient preferences, patient age, clinical comorbidities, and clinical probability of finding an APA (see Figure 10–6).

c. **Glucocorticoid-remediable aldosteronism—familial hyperaldosteronism type 1**—In the absence of glucocorticoid therapy, this mutation results in overproduction of aldosterone and the hybrid steroids 18-hydroxycortisol and 18-oxycortisol, which can be measured in the urine to make the diagnosis. Genetic testing is a sensitive and specific means of diagnosing GRA and obviates the need to measure the urinary levels of 18-oxocortisol and 18-hydroxycortisol or to perform dexamethasone suppression testing. Genetic testing for GRA should be considered for primary aldosteronism patients with a family history of primary aldosteronism or onset of primary aldosteronism at a young age (eg, <20 years), or in primary aldosteronism patients who have a family history of strokes at a young age. Cerebrovascular complications (eg, hemorrhagic stroke) associated with intracranial aneurysms affect approximately 20% of all patients with GRA—a frequency of cerebral aneurysm similar to that found in adult polycystic kidney disease.

d. **Familial hyperaldosteronism type 2**—FH type 2 is autosomal dominant and may be monogenic. All hypertensive first degree relatives of patients with primary aldosteronism should be advised to have case detection testing for primary aldosteronism.

e. **Familial hyperaldosteronism type 3**—FH type 3 typically presents in childhood or adolescence with hypertension, which may be mild or severe. When hypertension is severe, spontaneous hypokalemia is more likely to be present. FH-3 is caused by germline mutations in the potassium channel *KCNJ5*. Germline mutation testing is available and should be considered in patients with familial hyperaldosteronism and in patients who present with primary aldosteronism in childhood or adolescence.

Treatment

The treatment goal is to prevent the morbidity and mortality associated with hypertension, hypokalemia, and cardiovascular damage. The cause of the primary aldosteronism helps to determine the appropriate treatment. Normalization of blood pressure should not be the only goal in managing a patient who has primary aldosteronism. In addition to the kidney and colon, mineralocorticoid receptors occur in the heart, brain, and blood vessels. Excessive secretion of aldosterone is associated with increased risk of cardiovascular disease and morbidity. Therefore, normalization of circulating aldosterone or mineralocorticoid receptor blockade should be part of the management plan for all patients with primary aldosteronism. However, clinicians must understand that most patients with longstanding primary aldosteronism have some degree of renal insufficiency that is masked by the glomerular hyperfiltration associated with aldosterone excess. The true degree of renal insufficiency may only become evident after effective pharmacologic or surgical therapy.

A. Surgical treatment of aldosterone-producing adenoma and unilateral hyperplasia Unilateral laparoscopic adrenalectomy is an excellent treatment option for patients with APA or PAH (unilateral hyperplasia). Although blood pressure control improves in nearly 100% of patients postoperatively, average long-term cure rates of hypertension after unilateral

adrenalectomy for APA range from 30% to 60%. Persistent hypertension following adrenalectomy is correlated directly with having more than one first-degree relative with hypertension, use of more than two antihypertensive agents preoperatively, older age, increased serum creatinine level, and duration of hypertension.

Laparoscopic adrenalectomy is the preferred surgical approach and is associated with shorter hospital stays and less long-term morbidity than the open approach. Because APAs are small and may be multiple, the entire adrenal gland should be removed. To decrease the surgical risk, hypokalemia should be corrected with potassium supplements and/or a mineralocorticoid receptor antagonist preoperatively. The mineralocorticoid receptor antagonist and potassium supplements should be discontinued postoperatively. PAC should be measured 1 to 2 days after the operation to confirm a biochemical cure. Serum potassium levels should be monitored weekly for 4 weeks after surgery and a generous sodium diet should be followed to avoid the hyperkalemia of hypoaldosteronism that may occur because of the chronic suppression of the renin-angiotensin-aldosterone axis. In approximately 5% of APA patients, clinically significant hyperkalemia may develop after surgery, and short-term fludrocortisone supplementation may be required. Typically, the component of hypertension that was associated with aldosterone excess resolves in 1 to 3 months postoperatively.

B. Pharmacologic treatment IHA should be treated medically. In addition, patients with APAs may be treated medically if the medical treatment includes mineralocorticoid receptor blockade. A sodium-restricted diet (<100 mEq of sodium per day), maintenance of ideal body weight, tobacco avoidance, and regular aerobic exercise contribute significantly to the success of pharmacologic treatment. No placebo-controlled randomized trials have evaluated the relative efficacy of drugs in the treatment of primary aldosteronism. Spironolactone, available as 25-, 50-, and 100-mg tablets, has been the drug of choice to treat primary aldosteronism for more than 4 decades. The initial dosage is 12.5 to 25 mg/d and is increased to 400 mg/d, if necessary, to achieve a high-normal serum potassium concentration without the aid of oral potassium chloride supplementation. Hypokalemia responds promptly, but hypertension may take as long as 4 to 8 weeks to correct. After several months of therapy, this dosage often can be decreased to as little as 25 to 50 mg/d; dosage titration is based on a goal serum potassium level in the high-normal range. Serum potassium and creatinine should be monitored frequently during the first 4 to 6 weeks of therapy (especially in patients with renal insufficiency or diabetes mellitus). Spironolactone increases the half-life of digoxin, and for patients taking this drug, the dosage may need to be adjusted when treatment with spironolactone is started. Concomitant therapy with salicylates should be avoided because they interfere with the tubular secretion of an active metabolite and decrease the effectiveness of spironolactone. Spironolactone is not selective for the mineralocorticoid receptor. For example, antagonism at the androgen receptor may result in painful gynecomastia, erectile dysfunction, and decreased libido in men; agonist activity at the progesterone receptor results in menstrual irregularity in women.

Eplerenone is a steroid-based antimineralocorticoid that acts as a competitive and selective mineralocorticoid receptor antagonist. It has a marked reduction in progestational and antiandrogenic actions compared with spironolactone. Eplerenone is available as 25- and 50-mg tablets. For primary aldosteronism, it is reasonable to start with a dose of 25 mg twice daily (twice daily because of the shorter half-life of eplerenone compared to spironolactone) and titrated upward for a target high-normal serum potassium concentration without the aid of potassium supplements. Potency studies with eplerenone show 50% less milligram per milligram potency when compared with spironolactone. As with spironolactone, it is important to follow blood pressure, serum potassium, and serum creatinine levels closely.

Patients with IHA frequently require a second antihypertensive agent to achieve adequate blood pressure control. Hypervolemia is a major reason for resistance to drug therapy, and low doses of a thiazide (eg, 12.5-50 mg of hydrochlorothiazide daily) or a related sulfonamide diuretic are effective in combination with the mineralocorticoid receptor antagonist. Because these agents often lead to further hypokalemia, serum potassium levels should be monitored.

Before initiating treatment, GRA should be confirmed with genetic testing. In the GRA patient, chronic treatment with physiologic doses of a glucocorticoid normalizes blood pressure and corrects hypokalemia. The clinician should avoid iatrogenic Cushing syndrome with excessive doses of glucocorticoids, especially with the use of dexamethasone in children. The smallest effective dose of shorter acting agents, such as prednisone or hydrocortisone, should be prescribed in relation to body surface area (eg, hydrocortisone, 10-12 mg/m^2/d). Target blood pressure in children should be guided by age-specific blood pressure percentiles. Children should be monitored by pediatricians with expertize in glucocorticoid therapy, with careful attention paid to preventing retardation of linear growth by overtreatment. Treatment with mineralocorticoid receptor antagonists in these patients may be just as effective and avoids the potential disruption of the hypothalamic-pituitary-adrenal axis and risk of iatrogenic side effects. In addition, glucocorticoid therapy or mineralocorticoid receptor blockade may even have a role in normotensive GRA patients.

Patients with FH-3 should be treated with mineralocorticoid receptor blockade. Some of these patients have severe aldosterone excess, and the daily required dosage of spironolactone or eplerenone is not feasible lifelong, and bilateral adrenalectomy may need to be considered in selected cases.

C. Pregnancy   Primary aldosteronism is uncommon in pregnancy with fewer than 40 patients reported in the medical literature, and most patients have had APA. Primary aldosteronism can lead to intrauterine growth retardation, preterm delivery, intrauterine fetal demise, and placental abruption. Case detection testing for primary aldosteronism in the pregnant woman is the same as for nonpregnant patients: morning blood sample for the measurement of aldosterone and plasma renin activity or renin mass measurement. If spontaneous hypokalemia in present in the woman with high aldosterone and suppressed renin, confirmatory testing is not needed. In the normokalemic woman with a positive case detection test, confirmatory testing should be pursued. However, the captopril stimulation test is contraindicated in pregnancy, and the saline infusion test may not be well tolerated. One option is measurement of sodium and aldosterone in a 24-hour urine collection on an ambient sodium diet.

Subtype testing with abdominal MRI without gadolinium is the test of choice. Adrenal imaging with CT, iodocholesterol scintigraphy, and adrenal venous sampling should be avoided in pregnancy.

Primary aldosteronism in pregnancy is fascinating in that the degree of disease may be improved or aggravated by pregnancy. In some women with primary aldosteronism, the high blood levels of pregnancy-related progesterone are antagonistic at the mineralocorticoid receptor and partially block the action of aldosterone—these patients have an improvement in the manifestations of primary aldosteronism during pregnancy. In other pregnant women, increased expression of luteinizing hormone-chorionic gonadotropin receptor and gonadotropin-releasing hormone receptor have been documented in APAs and the degree of hyperaldosteronism is aggravated by the increased pregnancy-related blood levels of human chorionic gonadotropin.

The type of treatment for primary aldosteronism in pregnancy depends on how difficult it is to manage the hypertension and hypokalemia. If the patient is in the subset of patients who have a remission in the degree of primary aldosteronism, then surgery or treatment with a mineralocorticoid antagonist can be avoided until after delivery. However, if hypertension and hypokalemia are marked, then surgical and/or medical intervention is indicated. Unilateral laparoscopic adrenalectomy during the second trimester can be considered in those women with confirmed primary aldosteronism and a clear-cut unilateral adrenal macroadenoma (>10 mm).

Spironolactone crosses the placenta and is a United States Federal Drug Association (FDA) pregnancy category C drug because feminization of newborn male rats has been documented. However, there is only one human case in the medical literature where treatment with spironolactone in pregnancy led to ambiguous genitalia in a male infant—this occurred in a woman treated with spironolactone for polycystic ovarian disease pre-pregnancy and through the fifth week of gestation. Eplerenone is an FDA pregnancy category B drug. Therefore, for those pregnant women who will be managed medically, the hypertension should be treated with standard antihypertensive drugs approved for use during pregnancy. Hypokalemia, if present, should be treated with oral potassium supplements. For those patients with refractory hypertension and/or hypokalemia, the addition of eplerenone may be cautiously considered.

OTHER FORMS OF MINERALOCORTICOID EXCESS OR EFFECT

The medical disorders associated with excess mineralocorticoid effect from 11-deoxycorticosterone (DOC) and cortisol are listed in Table 10–2. These diagnoses should be considered when PAC and PRA are low in patients with hypertension and hypokalemia.

Hyperdeoxycorticosteronism

A. Congenital adrenal hyperplasia Congenital adrenal hyperplasia (CAH) is caused by enzymatic defects in adrenal steroidogenesis that result in deficient secretion of cortisol (see Chapter 14). The lack of inhibitory feedback by cortisol on the hypothalamus and pituitary produces an ACTH-driven buildup of cortisol precursors proximal to the enzymatic deficiency. A deficiency of 11β-hydroxylase (CYP11B) or 17α-hydroxylase (CYP17) causes hypertension and hypokalemia because of hypersecretion of the mineralocorticoid DOC. The mineralocorticoid effect of increased circulating levels of DOC also decreases PRA and aldosterone secretion. These defects are autosomal recessive in inheritance and typically are diagnosed in childhood. However, partial enzymatic defects have been shown to cause hypertension in adults.

a. **11β-Hydroxylase deficiency**—Approximately 5% of all cases of CAH are due to 11β-hydroxylase deficiency; the prevalence in Caucasians is 1 in 100,000. More than 40 mutations have been described in *CYP11B1*, the gene encoding 11β-hydroxylase. There is an increased prevalence among Sephardic Jews from Morocco, suggestive of a founder effect. The impaired conversion of DOC to corticosterone results in high levels of DOC, 11-deoxycortisol, and adrenal androgens (see Chapter 14). Females present in infancy or childhood with hypertension, hypokalemia, acne, hirsutism, and virilization. Males present with pseudoprecocious puberty. Approximately two-thirds of patients have mild to moderate hypertension. Markedly increased levels of DOC, 11-deoxycortisol, and adrenal androgens confirm the diagnosis. Glucocorticoid replacement normalizes the steroid abnormalities and hypertension.

b. **17α-Hydroxylase deficiency**—17α-Hydroxylase deficiency is a rare cause of CAH. 17α-Hydroxylase is essential for the synthesis of cortisol and gonadal hormones, and deficiency results in decreased production of cortisol and sex hormones. Genetic 46,XY males present with either pseudohermaphroditism or as phenotypic females, and 46,XX females present with primary amenorrhea. Therefore, a person with this form of CAH may not come to medical attention until puberty. Patients present with eunuchoid proportions and appearance. The biochemical findings include low concentrations of plasma adrenal androgens, plasma 17α-hydroxyprogesterone and cortisol. The plasma concentrations of DOC, corticosterone, and 18-hydroxycorticosterone (all produced in the zona fasciculata) are increased. Aldosterone (produced in zona glomerulosa) and PRA are suppressed. Although rare, there is an increased prevalence among Dutch Mennonites. As with 11β-hydroxylase deficiency, glucocorticoid replacement normalizes the steroid abnormalities and hypertension.

B. Deoxycorticosterone-producing tumor DOC-producing adrenal tumors are usually large and malignant. Some of them secrete androgens and estrogens in addition to DOC, which may cause virilization in women and feminization in men. A high level of plasma DOC or urinary tetrahydrodeoxycorticosterone and a large adrenal tumor seen on CT confirm the diagnosis. Aldosterone secretion in these patients is typically suppressed. Optimal treatment is complete surgical resection.

C. Primary cortisol resistance Increased cortisol secretion and plasma cortisol concentrations without evidence of Cushing syndrome are found in patients with primary cortisol resistance (or glucocorticoid resistance), a rare familial syndrome. Cortisol resistance leads to elevations in ACTH secretion which, in turn, increases adrenal steroid production. The syndrome is characterized by hypokalemic alkalosis, hypertension, increased plasma concentrations of DOC, as well as cortisol, and increased adrenal androgen secretion. The hypertension and hypokalemia are likely due to the combined effects of excess DOC and increased cortisol access to the mineralocorticoid receptor (high rates of cortisol production that overwhelm 11β-hydroxysteroid dehydrogenase type 2 [11β-HSD2] activity). Primary cortisol resistance is caused by defects in glucocorticoid receptors and assembly of the steroid–receptor complex. The treatment for the mineralocorticoid-dependent hypertension is blockade of the mineralocorticoid receptor with a mineralocorticoid receptor antagonist or suppression of ACTH secretion with dexamethasone.

Apparent Mineralocorticoid Excess Syndrome

Type 1 apparent mineralocorticoid excess (AME) is the result of impaired activity of the microsomal enzyme 11β-HSD2, which normally inactivates cortisol in the kidney by converting it to cortisone (Figure 10–9). Cortisol can be a potent mineralocorticoid, and as a result of the enzyme deficiency, high levels of
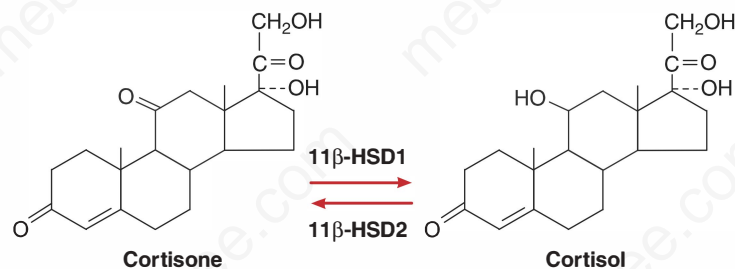
FIGURE 10–9 Interconversion of hormonally active cortisol to inactive cortisone is catalyzed by two isozymes of 11β-hydroxysteroid dehydrogenase (11β-HSD). 11β-HSD1 converts cortisone to cortisol and 11β-HSD2 inactivates cortisol to cortisone.

cortisol accumulate in the kidney. Thus, 11β-HSD2 normally excludes physiologic glucocorticoids from the nonselective mineralocorticoid receptor by converting them to the inactive 11-keto compound, cortisone. The characteristic abnormal urinary cortisol-cortisone metabolite profile seen in apparent mineralocorticoid excess reflects decreased 11β-HSD2 activity (ratio of cortisol to cortisone increased 10-fold from normal).

Decreased 11β-HSD2 activity may be hereditary or secondary to pharmacologic inhibition of enzyme activity by glycyrrhizic acid, the active principle of licorice root (*Glycyrrhiza glabra*) and some chewing tobaccos. The congenital form is a rare autosomal recessive disorder, and children present with low birth weight, failure to thrive, hypertension, polyuria and polydipsia, and poor growth. The clinical phenotype of patients with AME includes hypertension, hypokalemia, metabolic alkalosis, low PRA, low PAC, and normal plasma cortisol levels. The diagnosis is confirmed by demonstrating an abnormal ratio of cortisol to cortisone in a 24-hour urine collection. Treatment includes blockade of the mineralocorticoid receptor with a mineralocorticoid receptor antagonist or suppression of endogenous cortisol secretion with dexamethasone.

Carbenoxolone (18β-glycyrrhetinic acid sodium hemisuccinate) is a glycyrrhizic acid metabolite that is used in some countries to treat gastroesophageal reflux disorder. Carbenoxolone acts by protecting the gastrointestinal mucosal barrier from acid-pepsin attack and increasing mucosal mucin production. However, carbenoxolone is also an inhibitor of 11β-HSD2 and side effects include sodium retention, hypokalemic alkalosis, suppressed plasma renin, and hypertension.

Type 2 AME caused by ectopic ACTH secretion, seen in patients with Cushing syndrome, is related to the high rates of cortisol production that overwhelm 11β-HSD2 activity. DOC levels may also be increased in severe ACTH-dependent Cushing syndrome and contribute to the hypertension and hypokalemia in this disorder.

Liddle Syndrome—Abnormal Renal Tubular Ionic Transport

In 1963, Liddle described an autosomal dominant renal disorder that appeared to be primary aldosteronism with hypertension, hypokalemia, and inappropriate kaliuresis. However, PAC and PRA were very low in patients with Liddle syndrome; thus, another name for this disorder is "pseudoaldosteronism." Liddle syndrome is caused by mutations in the β or γ subunits of the amiloride-sensitive epithelial sodium channel—resulting in enhanced activity of the epithelial sodium channel with increased sodium reabsorption, potassium wasting, hypertension, and hypokalemia. Clinical genetic testing is available (www.genetests.org). As would be predicted, amiloride and triamterene are very effective agents to treat the hypertension and hypokalemia. However, spironolactone is ineffective in these patients. Liddle syndrome can easily be distinguished from AME based on a good clinical response to amiloride or triamterene, lack of efficacy of spironolactone and dexamethasone, and normal 24-hour urine cortisone-cortisol ratio.

Hypertension Exacerbated by Pregnancy

Hypertension exacerbated by pregnancy is a rare autosomal dominant disorder found in women with early-onset hypertension with suppressed levels of aldosterone and renin. During pregnancy, both the hypertension and hypokalemia are severely exacerbated. These patients have an activating mutation in the gene encoding the mineralocorticoid receptor which allows progesterone and other mineralocorticoid antagonists to become agonists.

OTHER ENDOCRINE DISORDERS ASSOCIATED WITH HYPERTENSION

Cushing Syndrome

Hypertension occurs in 75% to 80% of patients with Cushing syndrome (see Chapter 9). Most patients with endogenous Cushing syndrome have ACTH-dependent disease due to a corticotroph pituitary adenoma (see Table 10–2). Ectopic ACTH secretion is the second most common cause of endogenous Cushing syndrome. The ACTH-independent causes (eg, adrenal cortisol-secreting adenoma or carcinoma) of Cushing syndrome are less common. Bilateral macronodular hyperplasia (BMAH) is associated with massive nodular hyperplasia of both adrenal glands. The pathophysiology of BMAH includes paracrine production of ACTH, ectopic expression of G-protein-coupled receptors, and germline mutations in the tumor suppressor gene armadillo repeat containing 5 (*ARMC5*) (see Chapter 9). ACTH-independent primary pigmented nodular adrenal disease (PPNAD) is a bilateral form of micronodular adrenal hyperplasia that leads to Cushing syndrome. Germline mutations in *PRKAR1A* cause PPNAD in the setting of Carney complex (see Chapter 9). The mechanisms of hypertension in the setting of Cushing syndrome include increased production of DOC, enhanced pressor sensitivity to endogenous vasoconstrictors (eg, epinephrine and angiotensin II), increased cardiac output, activation of the renin-angiotensin system by increasing the hepatic production of angiotensinogen, and overload of the cortisol inactivation system with stimulation of the mineralocorticoid receptor. The source of excess glucocorticoids may be exogenous (iatrogenic) or endogenous. Mineralocorticoid production is usually normal in endogenous Cushing syndrome; aldosterone and renin levels are usually normal and DOC levels are normal or mildly increased. In patients with adrenal carcinoma, DOC and aldosterone may be elevated.

The case-detection studies for endogenous cortisol excess include: (a) measurement of free cortisol in a 24-hour urine collection; (b) midnight salivary cortisol measurement; and (c) 1-mg overnight dexamethasone suppression test. Further studies to confirm Cushing syndrome and to determine the cause of the cortisol excess state are outlined in Chapter 9.

The hypertension associated with Cushing syndrome should be treated until a surgical cure is obtained. Mineralocorticoid receptor antagonists, at dosages used to treat primary aldosteronism, are effective in reversing the hypokalemia. Second-step agents (eg, thiazide diuretics) may be added for optimal control of blood pressure.

The hypertension associated with the hypercortisolism usually resolves over several weeks after a surgical cure, and antihypertensive agents can be tapered and withdrawn.

Thyroid Dysfunction

A. Hyperthyroidism   When excessive amounts of circulating thyroid hormones interact with thyroid hormone receptors on peripheral tissues, both metabolic activity and sensitivity to circulating catecholamines increase. Thyrotoxic patients usually have tachycardia, high cardiac output, increased stroke volume, decreased peripheral vascular resistance, and increased systolic blood pressure. The initial management of patients with hypertension who have hyperthyroidism includes a β-adrenergic blocker to treat hypertension, tachycardia, and tremor. The definitive treatment of hyperthyroidism is cause specific (see Chapter 7).

B. Hypothyroidism   The frequency of hypertension, usually diastolic, is increased threefold in hypothyroid patients and may account for as much as 1% of cases of diastolic hypertension in the population. The mechanisms for the elevation in blood pressure include increased systemic vascular resistance and extracellular volume expansion. Treatment of thyroid hormone deficiency decreases blood pressure in most patients with hypertension and normalizes blood pressure in one-third of them. Synthetic levothyroxine is the treatment of choice for hypothyroidism (see Chapter 7).

Acromegaly

Chronic growth hormone (GH) excess from a GH-producing pituitary tumor results in the clinical syndrome of acromegaly. The effects of chronic excess of GH include acral and soft tissue overgrowth, progressive dental malocclusion, degenerative arthritis related to chondral and synovial tissue overgrowth within joints, low-pitched sonorous voice, excessive sweating and oily skin, perineural hypertrophy leading to nerve entrapment (eg, carpal tunnel syndrome), cardiac dysfunction, and hypertension (see Chapter 4). Hypertension occurs in 20% to 40% of the patients and is associated with sodium retention and extracellular volume expansion. Pituitary surgery is the treatment of choice; if necessary, it is supplemented with medical therapy or irradiation or both. The hypertension of acromegaly is treated most effectively by eliminating GH excess. If a surgical cure is not possible, the hypertension usually responds well to diuretic therapy.

REFERENCES

Primary Aldosteronism

Azizan EA, Brown MJ. Novel genetic determinants of adrenal aldosterone regulation. *Curr Opin Endocrinol Diabetes Obes.* 2016;23:209. [PMID: 26992195]

Funder JW, Carey RM, Mantero F, et al. The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab.* 2016;101(5):1889-916. [PMID: 26934393]

Kempers MEJ, Lenders JWM, van Outheusden L, et al. Systematic review: diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. *Ann Intern Med.* 2009;151:329. [PMID: 19721021]

Korah HE, Scholl UI. An update on familial hyperaldosteronism. *Horm Metab Res.* 2015;47:941. [PMID: 26445452]

Lim V, Guo Q, Grant CS, et al. Accuracy of adrenal imaging and adrenal venous sampling in predicting surgical cure of primary aldosteronism. *J Clin Endocrinol Metab.* 2014;99:2712. [PMID: 24796926]

Muth A, Ragnarsson O, Johannsson G, Wängberg B. Systematic review of surgery and outcomes in patients with primary aldosteronism. *Br J Surg.* 2015;102:307. [PMID: 25605481]

Reincke M, Rump LC, Quinkler M, et al. Risk factors associated with a low glomerular filtration rate in primary aldosteronism. *J Clin Endocrinol Metab.* 2009;94:869. [PMID: 19116235]

Stowasser M. Update in primary aldosteronism. *J Clin Endocrinol Metab.* 2015;100:1. [PMID: 25365316]

Teo AE, Garg S, Shaikh LH, et al. Pregnancy, primary aldosteronism, and adrenal CTNNB1 mutations. *N Engl J Med.* 2015;373:1429. [PMID: 26397949]

Young WF, Stanson AW. What are the keys to successful adrenal venous sampling (AVS) in patients with primary aldosteronism? *Clin Endocrinol (Oxf).* 2009;70:14. [PMID: 19128364]

Other Forms of Mineralocorticoid Excess or Effect

Charmandari E, Kino T, Chrousos GP. Primary generalized familial and sporadic glucocorticoid resistance (Chrousos syndrome) and hypersensitivity. *Endocr Dev.* 2013;24:67. [PMID: 23392096]

Geller DS, Farhi A, Pinkerton N, et al. Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science.* 2000;289:119. [PMID: 10884226]

Hanukoglu I, Hanukoglu A. Epithelial sodium channel (ENaC) family: phylogeny, structure-function, tissue distribution, and associated inherited diseases. *Gene.* 2016;579:95. [PMID: 26772908]

Hassan-Smith Z, Stewart PM. Inherited forms of mineralocorticoid hypertension. *Curr Opin Endocrinol Diabetes Obes.* 2011;18:177. [PMID: 21494136]

Merke DP, Bornstein SR. Congenital adrenal hyperplasia. *Lancet.* 2005;365:2125. [PMID: 15964450]

Razzaghy-Azar M, Yau M, Khattab A, New MI. Apparent mineralocorticoid excess and the long term treatment of genetic hypertension. *J Steroid Biochem Mol Biol.* 2017;165(Pt A):145. [PMID: 26892095]

Other Endocrine Disorders Associated with Hypertension

Berg C, Petersenn S, Lahner H, et al. Cardiovascular risk factors in patients with uncontrolled and long-term acromegaly: comparison with matched data from the general population and the effect of disease control. *J Clin Endocrinol Metab.* 2010;95:3648. [PMID: 20463098]

Danzi S, Klein I. Thyroid hormone and blood pressure regulation. *Curr Hypertens Rep.* 2003;5:513. [PMID: 14594573]

Isidori AM, Graziadio C, Paragliola RM, et al. The hypertension of Cushing's syndrome: controversies in the pathophysiology and focus on cardiovascular complications. *J Hypertens.* 2015;33(1):44. [PMID: 25415766]

Streeten DH, Anderson GH Jr, Howland T, et al. Effects of thyroid function on blood pressure. Recognition of hypothyroid hypertension. *Hypertension.* 1988;11:78. [PMID: 3338842]

Terzolo M, Matrella C, Boccuzzi A, et al. Twenty-four hour profile of blood pressure in patients with acromegaly. Correlation with demographic, clinical and hormonal features. *J Endocrinol Invest.* 1999;22:48. [PMID: 10090137]

This page intentionally left blank

Adrenal Medulla and Paraganglia

Paul A. Fitzgerald, MD

ACE	Angiotensin-converting enzyme		**MIBG**	Metaiodobenzylguanidine
ACTH	Adrenocorticotropic hormone		**NF-1**	Neurofibromatosis type 1
AM	Adrenomedullin		**NSE**	Neuron-specific enolase
APMO	Adnexal cystadenomas of probable mesonephric origin		**PGL**	Paraganglioma
			PHEO	Pheochromocytoma
ARDS	Acute respiratory distress syndrome		**PNMT**	Phenylethanolamine-*N*-methyltransferase
ATP	Adenosine triphosphate			
cAMP	Cyclic adenosine monophosphate		**PTHrP**	Parathyroid hormone–related peptide
CCB	Calcium channel blocker			
CgA	Chromogramin A		**RECIST**	Response evaluation criteria in solid tumors
CGRP	Calcitonin gene–related peptide			
COMT	Catecholamine-*O*-methyltransferase		**SDHA**	Succinate dehydrogenase subunit A
DBH	Dopamine-β-hydroxylase		**SDHB**	Succinate dehydrogenase subunit B
DHPG	Dihydroxyphenylglycol		**SDHC**	Succinate dehydrogenase subunit C
ECD	Electrochemical detection		**SDHD**	Succinate dehydrogenase subunit D
FDG	Fluorodeoxyglucose		**SDHx**	Succinate dehydrogenase subunits
HIF	Hypoxia-inducible factor		**SPECT**	Single photon emission computed tomography
HN-PGL	Head-neck paraganglioma			
HPLC	High-pressure liquid chromatography		**SRI**	Somatostatin receptor imaging
			VEGF	Vascular endothelial growth factor
IP₃	Inositol triphosphate		**VHL**	von Hippel-Lindau
MAO	Monoamine oxidase		**VIP**	Vasoactive intestinal polypeptide
MEN	Multiple endocrine neoplasia		**VMA**	Vanillylmandelic acid (3-methoxy-4-hydroxymandelic acid)
MHBA	3-methoxy-4-hydroxybenzylamine			

The adrenal medulla and paraganglia are part of the autonomic/sympathetic nervous system. The endocrine and nervous systems are alike in that they exert their actions by releasing hormones/neurotransmitters that bind to cell surface receptors in the target tissue, thereby inducing an effect.

Autonomic nerves are not under conscious control. They innervate the heart, adrenal medulla, vascular smooth muscle, and smooth muscle in visceral organs, thereby controlling cardiac rate and output, adrenal medullary secretion of catecholamines, blood pressure, the genitourinary tract, and intestinal motility. Autonomic nerves originate within the central nervous system and have two major divisions according to their anatomic locations:

1. **Parasympathetic** preganglionic nerves exit the central nervous system via the cranial nerves and sacral spinal nerves. They terminate in nonchromaffin paraganglia that are most numerous in the neck and associated with the glossopharyngeal and vagus nerves. These ganglia serve as chemoreceptors that are involved in the control of respiration. An important

paraganglion at the carotid bifurcation is known as the **carotid body**. They are also found along the jugular vein and in the jugulotympanic region. Head-neck paragangliomas are tumors that arise from these parasympathetic paraganglia.

2. **Sympathetic** preganglionic nerves exit the central nervous system via the thoracic and lumbar spinal nerves. The sympathetic nervous system coordinates the body's automatic *fight-flight* response by stimulating the adrenal medulla to secrete catecholamines and by directly stimulating cardiac output and blood flow to muscles while diverting blood flow away from visceral organs.

Sympathetic preganglionic nerves terminate mainly in paravertebral and prevertebral nerve ganglia where they secrete acetylcholine as their neurotransmitter; they are, therefore, known as *cholinergic* nerves. These nerve ganglia are collectively known as **paraganglia** and contain neuroendocrine cells that are similar to adrenal medullary cells on light microscopy by chromaffin and immunohistochemical staining. Paraganglia are also found in the mediastinum, particularly adjacent to the cardiac atria, and in the abdomen along the sympathetic nerve chains in paravertebral and prevertebral positions. Paraganglia are plentiful along the aorta, particularly around the celiac axis, adrenal glands, renal medullae, and aortic bifurcation (organ of Zuckerkandl). Paraganglia are also abundant in the pelvis, particularly adjacent to the bladder. Preganglionic nerves also terminate in the adrenal medulla, which is basically a sympathetic ganglion that is surrounded by adrenal cortex.

Sympathetic postganglionic nerve fibers originate from the paraganglia and run to the tissues being innervated. They secrete norepinephrine as their neurotransmitter at synaptic junctions. Adrenal medullary cells are basically modified postganglionic nerves that lack axons and secrete their neurotransmitter (mainly epinephrine) directly into the blood; thus, the bloodstream acts like a giant synapse, carrying epinephrine to receptors throughout the body. Synonyms for epinephrine and norepinephrine are *adrenalin* and *noradrenaline*, respectively (discussed later).

Although the adrenal medulla is not essential for survival, its secretion of epinephrine and other compounds helps maintain the body's homeostasis during stress. Investigations of the adrenal medulla and the sympathetic nervous system have led to the discovery of different catecholamine receptors and the production of a wide variety of sympathetic agonists and antagonists with diverse clinical applications.

Pheochromocytomas are tumors that arise from the adrenal medulla, whereas non–head-neck **paragangliomas** arise from extra-adrenal sympathetic ganglia. Pheochromocytomas can secrete excessive amounts of both epinephrine and norepinephrine, whereas most paragangliomas secrete only norepinephrine. The excessive secretion of catecholamines can result in a dangerous exaggeration of the stress response.

ANATOMY

Embryology (Figure 11–1)

The sympathetic nervous system arises in the fetus from the primitive cells of the neural crest (sympathogonia). At about the fifth
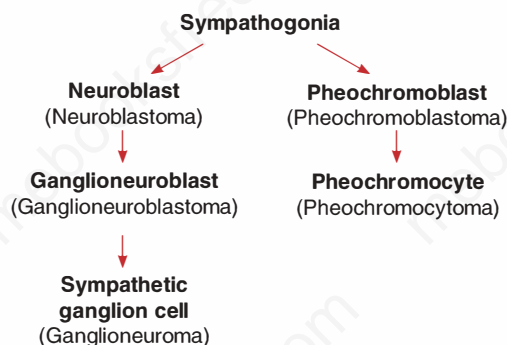
FIGURE 11–1 The embryonic development of adrenergic cells and tumors that develop from them (in parentheses). Sympathogonia are primitive cells derived from the neural crest. Neuroblasts are also called sympathoblasts; ganglion cells are the same as sympathocytes; and pheochromocytes are mature chromaffin cells.

week of gestation, these cells migrate from the spinal ganglia in the thoracic region to form the sympathetic chain posterior to the dorsal aorta. They then begin to migrate anteriorly to form the remaining ganglia.

At 6 weeks of gestation, groups of these primitive cells migrate along the central vein and enter the fetal adrenal cortex to form the adrenal medulla, which is detectable by the eighth week. The adrenal medulla at this time is composed of sympathogonia and pheochromoblasts, which then mature into pheochromocytes. The cells appear in rosette-like structures, with the more primitive cells occupying a central position. Storage granules can be found in these cells at 12 weeks. The adrenal medullas are very small and amorphous at birth but develop into recognizable adult form by the sixth month of postnatal life.

Pheochromoblasts and pheochromocytes also collect on both sides of the aorta to form the paraganglia. These cells collect principally at the origin of the mesenteric arteries and at the aortic bifurcation where they fuse anteriorly to form the organ of Zuckerkandl, which is quite prominent during the first year of life. Pheochromocytes (chromaffin cells) also are found scattered throughout the abdominal sympathetic plexi as well as in other parts of the sympathetic nervous system.

Gross Structure

The anatomic relationships between the adrenal medulla and the adrenal cortex differ in different species. In mammals, the medulla is surrounded by the adrenal cortex, and in humans, the adrenal medulla occupies a central position in the widest part of the gland, with only small portions extending into the narrower parts. The mass of adrenal medullary tissue in both adult adrenal glands averages about 1000 mg (about 15% of the total weight of both adrenal glands), although the proportions vary from individual to individual. There is no clear demarcation between cortex and medulla. A cuff of adrenal cortical cells usually surrounds the central vein within the adrenal medulla, and there may be islands of cortical cells elsewhere in the medulla.

Microscopic Structure

The **chromaffin cells** or **pheochromocytes** of the adrenal medulla are large ovoid columnar cells arranged in nests, alveoli, or cords around a rich network of capillaries and venous sinusoids that drain blood from the adrenal cortex. Pheochromocytes have large nuclei and a well-developed Golgi apparatus. Their cytoplasm contains large numbers of vesicles (granules) that measure 100 to 300 nm in diameter and appear similar to the neurosecretory granules seen in peripheral sympathetic nerves. Catecholamines (epinephrine and/or norepinephrine) comprise about 20% of the mass of neurosecretory vesicles. Vesicles containing norepinephrine appear darker than those containing epinephrine. The vesicles also contain proteins, lipids, and adenosine triphosphate (ATP), as well as chromogranins, neuropeptide Y, enkephalins, and proopiomelanocortin (along with related peptides such as adrenocorticotropic hormone [ACTH] and β-endorphin).

Nerve Supply

The cells of the adrenal medulla are innervated by preganglionic fibers of the sympathetic nervous system, which release acetylcholine and enkephalins at the synapses. Most of these fibers arise from a plexus in the capsule of the posterior surface of the gland and enter the adrenal glands in bundles of 30 to 50 fibers without synapsing. They follow the course of the blood vessels into the medulla without branching into the adrenal cortex. Some reach the wall of the central vein, where they synapse with small autonomic ganglia. However, most fibers end in relationship to the pheochromocytes.

Blood Supply

The adrenal gland is usually perfused by the superior, middle, and inferior adrenal branches of the inferior phrenic artery, directly from the aorta and from the renal arteries. On reaching the adrenal gland, these arteries branch to form a plexus under the capsule supplying the adrenal cortex. A few of these vessels, however, penetrate the cortex, passing directly to the medulla. The medulla is also nourished by branches of the arteries supplying the central vein and cuff of cortical tissue around the central vein. Capillary loops passing from the subcapsular plexus of the cortex also supply blood as they drain into the central vein. Most of the blood supply to adrenal medullary cells is via a portal vascular system that arises from the capillaries in the cortex. There is also a capillary network of lymphatics that drain into a plexus around the central vein.

Norepinephrine is converted to epinephrine by the enzyme phenylethanolamine-*N*-methyltransferase (PNMT). In mammals, the expression of PNMT is induced by cortisol. Chromaffin cells that produce epinephrine are exposed to higher concentrations of cortisol from capillaries draining adrenocortical cells, whereas chromaffin cells that produce norepinephrine are supplied by arteries that course directly to the adrenal medulla (see sections on biosynthesis and secretion, discussed later).

The central vein of the right adrenal is short and drains directly into the vena cava with about 5% having multiple veins. About 5% of right adrenal veins drain into the hepatic vein. For the left adrenal gland, the vein is somewhat longer and drains into the left renal vein.

HORMONES OF THE ADRENAL MEDULLA AND PARAGANGLIA

CATECHOLAMINES

Biosynthesis (Figure 11–2)

Catecholamines are molecules that have a catechol nucleus consisting of benzene with two hydroxyl side groups plus a side-chain amine. Catecholamines include dopamine, norepinephrine, and epinephrine (see Figure 11–2).

Epinephrine is the main hormone secreted by the normal adrenal medulla. The proportions of epinephrine and norepinephrine
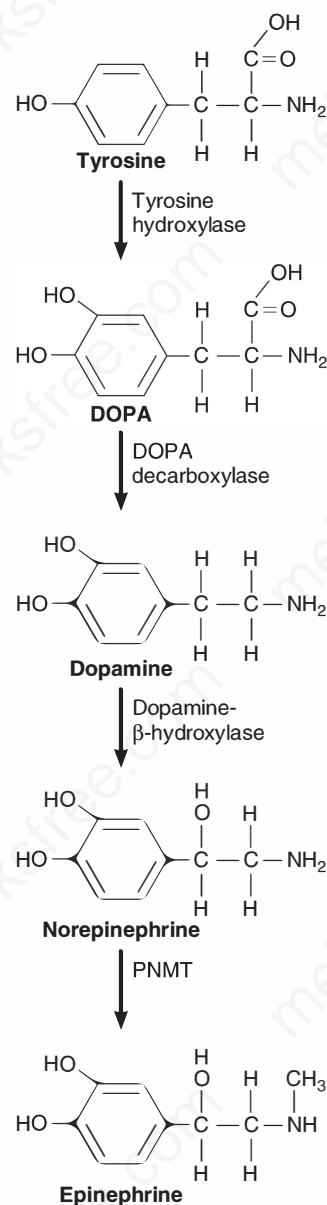
FIGURE 11–2 Biosynthesis of catecholamines (DOPA: L-dihydroxyphenylalanine; PNMT: phenylethanolamine-*N*-methyltransferase). Epinephrine is not produced in paraganglia.

found in the adrenal medulla vary with the species; in humans, the adrenal medulla contains 15% to 20% norepinephrine.

Norepinephrine is found primarily in the central nervous system and in peripheral sympathetic paraganglia and nerves. Dopamine is found in the adrenal medulla and in sympathetic neurons as a precursor to norepinephrine. It is present in high concentrations in the brain, in specialized neurons in the sympathetic ganglia, and in the carotid body, where it serves as a neurotransmitter. Dopamine is also present in the proximal renal tubule where it promotes sodium excretion and in the gastrointestinal tract where it serves a paracrine function.

Chromogranin A (CgA) is a peptide that is stored and released with catecholamines during exocytosis; catestatin is a fragment of the CgA prohormone that inhibits further catecholamine release by acting as an antagonist at the neuronal cholinergic receptor. CgA levels tend to be somewhat higher in patients with hypertension than in matched normotensive individuals. CgA has become a valuable tumor marker, particularly for patients with paragangliomas that are otherwise nonsecretory.

A. Conversion of tyrosine to DOPA   Tyrosine is the precursor to all catecholamines. Tyrosine is derived from ingested food and synthesized from phenylalanine in the liver. Tyrosine circulates at a concentration of 1 to 1.5 mg/dL of blood. It enters neurons and chromaffin cells by an active transport mechanism and is converted to L-3,4-dihydroxyphenylalanine (L-DOPA). Humans produce only the levo isomer, so L-DOPA is also referred to as simply DOPA. The reaction is catalyzed by *tyrosine hydroxylase*, which is transported via axonal flow to the nerve terminal. Tyrosine hydroxylase is the rate-limiting step in catecholamine synthesis. It is transcriptionally activated by acetylcholine through the nicotinic cholinergic receptor, which in turn activates protein kinase A via cAMP. Tyrosine hydroxylase activity may be inhibited by a variety of compounds; alpha methyltyrosine (metyrosine) is an inhibitor that may be used to reduce catecholamine secretion in patients with pheochromocytoma or paraganglioma.

B. Conversion of DOPA to dopamine   DOPA is converted to dopamine by the removal of its carboxyl group by the enzyme aromatic L-amino acid decarboxylase (DOPA decarboxylase). This enzyme is found in all tissues, with the highest concentrations in liver, kidney, vas deferens, and brain. In the brain, dopamine functions as an important neurotransmitter.

C. Conversion of dopamine to norepinephrine   Dopamine enters granulated storage vesicles where it is hydroxylated to norepinephrine by the enzyme dopamine-β-hydroxylase (DBH), which is found within the vesicle membrane. Norepinephrine is then actively transported into the vesicle by vesicular monoamine transferase (VMAT), which is located in the lipid bilayer of the vesicle wall. Norepinephrine constantly diffuses out of the vesicle, so it is best to consider it being concentrated in the vesicle, rather than being stored there. The granulated storage vesicle migrates to the cell surface and secretes its contents via exocytosis; both norepinephrine and DBH are released during exocytosis. After

secretion, most norepinephrine is avidly recycled back into the nerve. Normally, about 93% of circulating norepinephrine originates from diffusion out of nonadrenal sympathetic nerve cells and synapses, while only 7% originates from the adrenal medulla.

D. Conversion of norepinephrine to epinephrine   Norepinephrine constantly diffuses out from storage granules into the cytoplasm. In most cells of the adrenal medulla, cytoplasmic norepinephrine is converted to epinephrine, catalyzed by PNMT. Epinephrine can then return to the vesicle, diffuse from the cell, or undergo catabolism. High concentrations of cortisol enhance the expression of the gene encoding PNMT. Cortisol is present in high concentrations in most areas of the adrenal medulla due to venous blood flow from the adjacent adrenal cortex. This accounts for the fact that in the normal human adrenal medulla, about 80% of the catecholamine content is epinephrine and only 20% is norepinephrine. The administration of exogenous corticosteroids increases adrenal medullary PNMT expression, despite suppression of ACTH and endogenous adrenal cortisol production

Paragangliomas rarely secrete epinephrine, because they do not have high local concentrations of cortisol to induce PNMT expression. Pheochromocytomas secrete epinephrine and norepinephrine in various ratios. Interestingly, after adrenalectomy, recurrent pheochromocytoma can secrete epinephrine without any surrounding adrenal cortex, indicating that the induction of PNMT may continue in subsequent metastatic cell lineages without high local levels of cortisol. After both adrenal glands are resected, serum epinephrine levels abruptly fall to nearly zero, whereas norepinephrine concentrations decline minimally, since most circulating norepinephrine originates from leakage out of sympathetic nerves and synapses.

The enzyme PNMT is found in many tissues outside the adrenal medulla. PNMT has been found in the lungs, kidneys, pancreas, and cancer cells. Therefore, nonadrenal tissue is capable of synthesizing epinephrine if norepinephrine is available as a substrate. However, nonadrenal production of epinephrine usually contributes minimally to circulating levels. PNMT is found in human lung, and glucocorticoids increase its expression, thereby increasing lung epinephrine. This may possibly contribute to the bronchial dilation that is induced by glucocorticoids (inhaled or systemic). PNMT activity is also found in red blood cells, where its activity is increased in hyperthyroidism and decreased in hypothyroidism. Renal PNMT activity is such that normally up to one-half of the epinephrine found in normal urine may be the product of renal conversion from norepinephrine.

In the adrenal medulla, catecholamine stores can be depleted during prolonged hypoglycemia. Biosynthesis of catecholamines increases during nerve stimulation by activation of tyrosine hydroxylase. Prolonged stimulation leads to the induction of increased amounts of this rate-limiting enzyme.

Storage of Catecholamines

Catecholamines are not truly stored in intracellular vesicles. Rather, they exist in a dynamic equilibrium between the vesicle

and the surrounding cytoplasm. Catecholamines continuously leak out of the vesicles into the cytoplasm but reenter and become concentrated in the vesicles by the action of two vesicular mono-amine transporters: VMAT-1 and VMAT-2. Both are expressed in the adrenal medulla, but only VMAT-2 is expressed in sympathetic nerves.

In the adrenal medulla, the percentage of stored epinephrine (vs total catecholamines) varies widely among different species. In the human adrenal medulla, epinephrine comprises about 80% of stored catecholamines. Epinephrine and norepinephrine are stored in different cells. Morphologic differences in epinephrine and norepinephrine storage vesicles are visible on light microscopy after proper staining. Cells that store predominantly epinephrine tend to be adjacent to intramedullary sinusoidal vessels that bathe the cells in blood from the adrenal cortex; such blood contains cortisol that induces the enzyme PNMT that catalyzes the conversion of norepinephrine to epinephrine in the cytoplasm such that these vesicles contain predominantly epinephrine. Adrenal medullary cells that contain predominantly norepinephrine tend to be farther away from such vessels.

Norepinephrine is the neurotransmitter that is secreted at the synapses of postganglionic sympathetic nerves in organs that have rich autonomic innervations: the heart, salivary glands, vascular smooth muscle, liver, spleen, kidneys, and muscles. A single sympathetic nerve cell may have up to 25,000 synaptic bulges along the length of its axon; each synapse synthesizes norepinephrine and stores it in adrenergic synaptic vesicles adjacent to target cells.

Catecholamine biosynthesis is coupled to secretion, so that the stores of norepinephrine at the nerve endings remain relatively unchanged even in the presence of marked nerve activity. The adrenal medulla contains catecholamines at a concentration of about 0.5 mg/g; the spleen, vas deferens, brain, spinal cord, and heart contain 1 to 5 μg/g; and the liver, gut, and skeletal muscle contain 0.1 to 0.5 μg/g. Catecholamines are stored in electron-dense vesicles approximately 1 μm in diameter that contain catecholamines and ATP in a 4:1 molar ratio, as well as several neuropeptides, calcium, magnesium, and water-soluble proteins called **chromogranins** (discussed earlier). The interior surface of the membrane contains DBH and ATPase. The Mg^{2+}-dependent ATPase facilitates the uptake and inhibits the release of catecholamines by the granules. Adrenal medullary granules appear to contain and release a number of active peptides including adrenomedullin, ACTH, vasoactive intestinal peptide (VIP), chromogranins, and enkephalins. The peptides derived from chromogranins are physiologically active and may modulate catecholamine release.

Secretion of Catecholamines (Figure 11–3 and Table 11–1)

Preganglionic nerve fibers terminate at adrenal medullary cells, where they release acetylcholine at synapses and stimulate adrenal medullary receptors. Activation of the receptors depolarizes the cell membranes and causes an influx of calcium ions. The increased intracellular concentration of calcium triggers exocytosis of the neurosecretory vesicles with consequent release of their contents:

catecholamines, chromogranins, and soluble DBH. Vesicular membrane-bound DBH is not released during exocytosis.

Adrenal medulla.   In the adrenal medulla, exocytosis of the neurosecretory vesicles releases epinephrine and norepinephrine into the circulation. In normals, about 95% of circulating epinephrine is derived from the adrenal medulla, whereas adrenal medullary secretion accounts for only about 7% of circulating norepinephrine. Catecholamines also diffuse out of the neurosecretory vesicles into the cytoplasm. Catecholamines in the cytoplasm may reenter a vesicle via VMAT or may be metabolized by COMT into catecholamine metabolites: metanephrine and normetanephrine, which gradually diffuse out of the cell and enter the circulation.

Adrenal medullary catecholamine secretion increases with exercise, angina pectoris, myocardial infarction, hemorrhage, ether anesthesia, surgery, hypoglycemia, anoxia and asphyxia, and many other stressful stimuli. The rate of secretion of epinephrine increases more than that of norepinephrine in the presence of hypoglycemia and most other stimuli. However, during hypoxia, the adrenal medullary cells preferentially release norepinephrine. Such release is mainly not by exocytosis but rather by leakage of norepinephrine from neurosecretory vesicles into the cytoplasm and out of the cell into the circulation. The preferential release of norepinephrine may be because the adrenal medullary cells that concentrate norepinephrine are located farther from blood vessels and may be more susceptible to hypoxic injury (see Storage of Catecholamines, discussed earlier).

Sympathetic neurons.   In sympathetic neurons, exocytosis of the neurosecretory vesicle releases norepinephrine into the synapse by exocytosis. Norepinephrine that has been secreted into a synapse is either reabsorbed via norepinephrine transporters (NETs) or leaks out of the synapse into the circulation. Such synaptic leakage accounts for about 93% of circulating norepinephrine in normal individuals.

Metabolism and Excretion of Catecholamines (Figures 11–4 and 11–5)

A. Catecholamine metabolism in cells of origin

Adrenal medulla.   In the adrenal medulla, catecholamines are metabolized to metanephrines primarily by membrane-bound catecholamine-*O*-methyltransferase (COMT): epinephrine to metanephrine and norepinephrine to normetanephrine. The metanephrine metabolites subsequently leak into the circulation. Normally, about 91% of circulating metanephrine is produced by the adrenal medulla, while the remaining 9% is derived from the metabolism of circulating epinephrine to metanephrine by the enzyme COMT found in the liver and other non-adrenal tissues. Similarly, in normals, about 23% of circulating normetanephrine is produced by the adrenal medulla, while 77% is derived from the metabolism of circulating norepinephrine to normetanephrine by the enzyme COMT found in the liver and other non-adrenal tissues (see Figure 11–5).
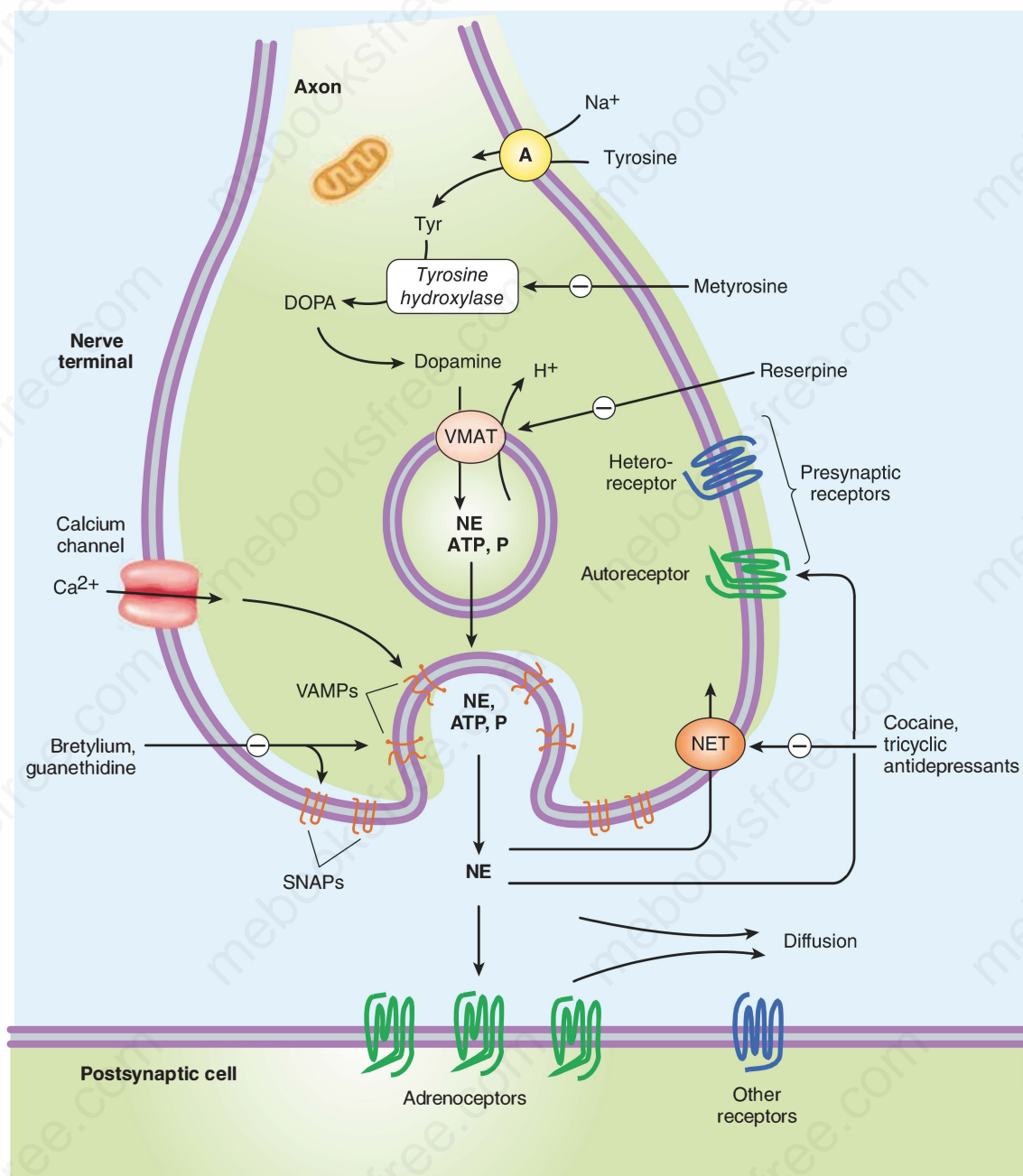
FIGURE 11–3 Schematic diagram of a noradrenergic junction of the peripheral sympathetic nervous system (not to scale). The neural axons form synaptic junctions with cells in target organs. The neurotransmitter is norepinephrine that has been either synthesized de novo from tyrosine or reabsorbed by a norepinephrine transporter (NET). Tyrosine is transported into the noradrenergic endings or varicosities by a sodium-dependent carrier (A). In the cytoplasm, tyrosine hydroxylase converts tyrosine to DOPA that is converted by DOPA decarboxylase to dopamine. Dopamine, as well as norepinephrine (NE) is transported into the vesicles by vesicular monoamine transferase (VMAT). Dopamine is converted to NE in the vesicle by DBH. Exocytosis of the neurosecretory vesicle with release of transmitter occurs when an action potential opens voltage-sensitive calcium channels and increases intracellular calcium. Fusion of vesicles with the surface membrane results in expulsion of norepinephrine, cotransmitters, and DBH. After release, norepinephrine diffuses out of the cleft or is transported into the cytoplasm of the nerve itself (uptake 1) or transported into the postjunctional target cell (uptake 2). Regulatory receptors are present on the presynaptic terminal (SNAPs, synaptosome-associated proteins; VAMPs, vesicle-associated membrane protein). **Not shown:** In adrenal medullary and pheochromocytoma cells, norepinephrine constantly diffuses from the vesicle into the cytoplasm where it may be converted to epinephrine by phenylethanolamine-*N*-methyltransferase (PNMT). Alternatively, cytoplasmic membrane-bound catecholamine-*O*-methyltransferase (COMT) can metabolize epinephrine and norepinephrine directly into metanephrine and normetanephrine, respectively, which are then released. (Reproduced with permission from Katzung BJ. *Basic & Clinical Pharmacology*. 11th ed. New York: McGraw-Hill Education; 2009.)

TABLE 11–1 Range of plasma catecholamine levels observed in healthy subjects and patients without pheochromocytoma or paraganglioma.

	Norepinephrine	Epinephrine
Healthy Subjects		
Basal (supine)	112-658 pg/mL (0.7-3.9 nmol/L)	25-50 pg/mL (0.1-0.3 nmol/L)
Ambulatory	217-1109 pg/mL (1.3-6.5 nmol/L)	30-95 pg/mL (0.2-0.5 nmol/L)
Exercise	800-4000 pg/mL (4.7-23.6 nmol/L)	100-1000 pg/mL (0.5-5 nmol/L)
Symptomatic hypoglycemia	200-1000 pg/mL (1.2-5.9 nmol/L)	1000-5000 pg/mL (5-25 nmol/L)
Patients		
Hypertension	200-500 pg/mL (1.2-3 nmol/L)	20-100 pg/mL (0.1-0.6 nmol/L)
Surgery	500-2000 pg/mL (3-12 nmol/L)	199-500 pg/mL (0.5-3 nmol/L)
Myocardial infarction	1000-2000 pg/mL (6-12 nmol/L)	800-5000 pg/mL (4-25 nmol/L)

Sympathetic nerves. Catecholamines that are released at the synapse bind to their receptors with relatively low affinity and dissociate rapidly (see Figure 11–4). About 10% of norepinephrine escapes from synapses into the systemic circulation (discussed later). About 90% of synaptic norepinephrine is reabsorbed by the nerves from which they were released or by the target cells.

Cytoplasmic norepinephrine is metabolized to 3,4-dihydroxyphenylglycol (DHPG) at the outer mitochondrial membrane by the enzyme monoamine oxidase (MAO) that regulates the norepinephrine content of neurons. Sympathetic nerves do not contain the enzyme COMT. DHPG leaks from the neuron and is absorbed by the liver and other tissues that contain the enzyme COMT that converts DHPG to 3-methoxy-4-hydroxyphenylglycol (MHPG), with subsequent conversion to vanillylmandelic acid (VMA) (discussed later) that is excreted in the urine. In normals, most urinary VMA is derived from DHPG that leaks from sympathetic nerves.

Sympathetic nerves in abdominal organs, particularly the intestines and pancreas, account for nearly half the body's production of norepinephrine. Intestinal cells contain sulfatases that conjugate norepinephrine to norepinephrine-SO_4 that enters the portal vein and bypasses the liver to be excreted in the urine. Intestinal cells also conjugate normetanephrine to normetanephrine-SO_4.

Pheochromocytomas. In pheochromocytomas (PHEOs), catecholamines that leak out of the neurosecretory vesicles are methylated by membrane-bound COMT: epinephrine to metanephrine and norepinephrine to normetanephrine. These metanephrine metabolites continuously leak directly into the circulation, in contrast to catecholamines that are secreted intermittently. The PHEO/PGL's continuous excretion of metanephrines accounts for the superior sensitivity of plasma or urine-fractionated free metanephrine testing in the diagnosis of
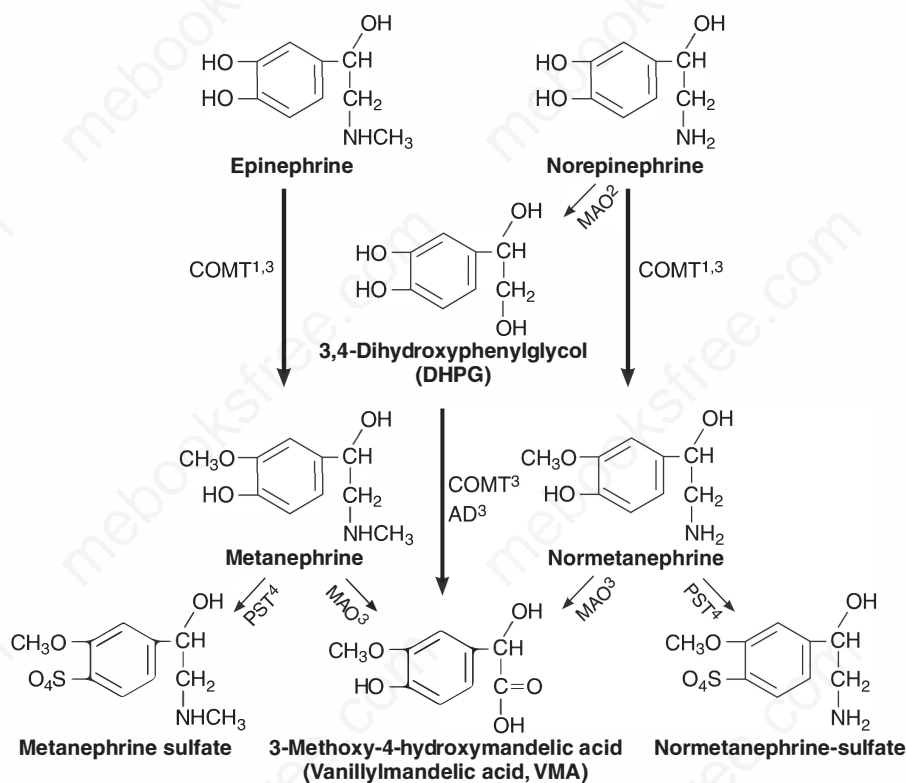
FIGURE 11–4 Metabolism of catecholamines by catechol-*O*-methyltransferase (COMT), monoamine oxidase (MAO), aldehyde dehydrogenase (AD), and phenol-sulfotransferase (PST). (1) Adrenal medulla or PHEO; (2) sympathetic nerves; (3) liver and kidneys; (4) GI, platelets, lungs.
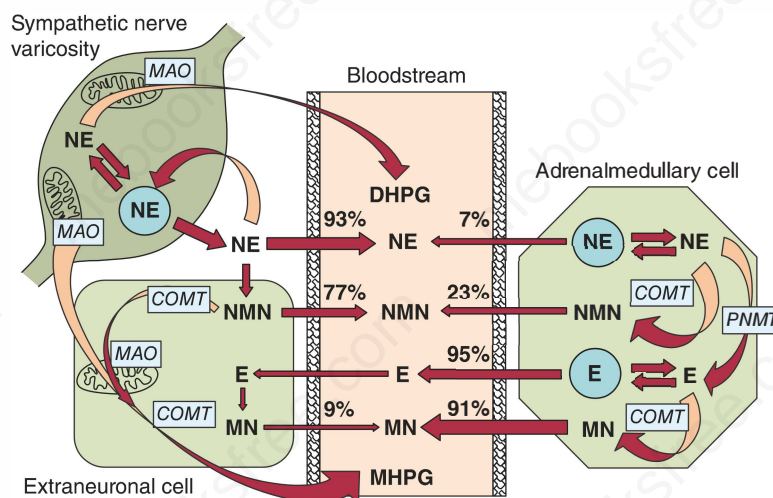
FIGURE 11–5 Location of catecholamine metabolism by normal adrenal medullary cells, sympathetic nerve cells, and extraneuronal cells (liver, kidneys) and their relative contributions to plasma levels. Most metabolism of norepinephrine (NE) in sympathetic nerves occurs from NE that has leaked from storage vesicles and undergoes deamination, in the sympathoneuronal cytoplasm that is catalyzed by monoamine oxidase (MAO) to form DHPG that leaks into the circulation. Similarly, in adrenal medullary chromaffin cells and chromaffin tumors (PHEO and PGL), metabolism occurs within cells from NE and epinephrine (E) leaking from storage vesicles into the cytoplasm, where the presence of catechol-amine-*O*-methyltransferase (COMT) leads to formation of normetanephrine (NMN) and metanephrine (MN), respectively, which leak into the circulation. COMT is not present in sympathetic nerves, but it is present in extraneuronal cells (liver, kidney), where it metabolizes DHPG to MHPG, the main precursor of vanilyl mandelic acid (VMA) that is excreted in the urine. The COMT in extraneuronal tissues is also responsible for production of a small proportion of circulating MN (9%), derived from E metabolized after release from chromaffin cells, as well as a larger proportion of circulating NMN (77%) derived mainly from NE released by sympathetic nerves. (Reproduced with permission from Eisenhofer G, Peitzsch M. Laboratory evaluation of pheochromocytoma and paraganglioma. *Clin Chem*. 2014 Dec;60(12):1486-1499.)

PHEO. In patients with PHEO, over 90% of circulating total metanephrines are derived from the tumor itself, rather than from peripheral metabolism.

B. Circulating catecholamine uptake and metabolism

Following release into the circulation, catecholamines bind to albumin and other proteins with low affinity and high capacity. Catecholamines are quickly removed from the bloodstream and have a circulating half-life of less than 2 minutes. Although some free catecholamines are excreted directly into the urine, most are actively transported from the circulation into other cells where they are metabolized. Organic cation transporters (OCTs) are largely responsible for removing catecholamines from the circulation; they are expressed in many tissues, particularly the liver. Following active transport into a cell, catecholamines and their immediate metabolites undergo further metabolism. The main extra-adrenal/neuronal catecholamine metabolic pathway is initially via the enzyme MAO that converts both epinephrine and norepinephrine to DHPG. DHPG is further metabolized by soluble COMT to MHPG that is converted by aldehyde dehydrogenase (AD) to VMA that is excreted in the urine.

Catecholamines, metanephrines, and MHPG also undergo conjugation of their phenolic hydroxyl group with sulfate or glucuronide; this reaction occurs mainly in gastrointestinal cells. To a lesser extent, catecholamines are also converted to metanephrines via a soluble form of COMT, an enzyme found in most tissues, especially blood cells, liver, kidney, and vascular smooth muscle.

The soluble form of COMT is less active than the membrane-bound form of COMT found in the adrenal medulla and pheochromocytomas.

Catecholamines, their metabolites, and conjugates are excreted in the urine. Normally, the proportions of urine catecholamines and metabolites are approximately 50% metanephrines, 35% VMA, 10% conjugated catecholamines and other metabolites, and less than 5% free catecholamines.

CATECHOLAMINE (ADRENERGIC) RECEPTORS (TABLES 11–2 AND 11–3; FIGURES 11–6 AND 11–7)

The adrenergic receptors are G protein–coupled receptors activated by norepinephrine and epinephrine. The finding that most cells in the body have adrenergic receptors has led to an appreciation of the important regulatory role of the peripheral sympathetic nervous system and circulating catecholamines.

Adrenergic receptors were originally classified into two groups: α and β. This classification has been expanded as subtypes of the α and β receptors have been discovered and includes α_{1A}, α_{1B}, α_{1C}, α_{2A}, α_{2B}, α_{2C} as well as β_1 and β_2, and β_3. There are five known dopamine receptor subtypes that are classified as either delta-1 type receptors or delta-2 type receptors. A different gene encodes each receptor subtype. Adrenergic receptors are variably distributed in the central nervous system and peripheral tissues.

TABLE 11–2 Catecholamine receptor types, subtypes, and cellular effects.

Receptor	Relative Agonist Potency	Effects
Alpha₁ Type Alpha$_{1A}$ Alpha$_{1B}$ Alpha$_{1D}$	Norepinephrine > epinephrine; midodrine	↑ IP$_3$, DAG common to all
Alpha₂ Type Alpha$_{2A}$	Norepinephrine > epinephrine; clonidine	↓ cAMP; ↑K$^+$ channels; ↓ Ca^{2+} channels
Alpha$_{2B}$	Norepinephrine > epinephrine	↓ cAMP; ↓Ca^{2+} channels
Alpha$_{2C}$		↓ cAMP
Beta Type Beta$_1$	Epinephrine = norepinephrine; dobutamine	↑ cAMP
Beta$_2$	Epinephrine > norepinephrine; albuterol	↑ cAMP
Beta$_3$	Norepinephrine > epinephrine	↑ cAMP
Delta-1 Type D$_1$ & D$_5$	Dopamine	↑ cAMP
Delta-2 Type D$_2$ & D$_3$ & D$_4$	Dopamine, bromocriptine, cabergoline	↓ cAMP

Data from Katzung BJ. *Basic & Clinical Pharmacology*. 13th ed. New York: McGraw-Hill Education; 2015.

Norepinephrine and epinephrine are roughly equipotent in activating α and β₁ receptors. Epinephrine is much more potent in activating β₂ receptors and norepinephrine is more potent in activating β₃ receptors (Table 11–2). The physiologic effects mediated by them are summarized in Table 11–3.

Although the adrenergic receptor subtypes exhibit significant amino acid homology, differences in the fifth and sixth segments determine the specificity of agonist binding. Differences in the fifth and seventh segments determine which of the guanylyl nucleotide binding proteins (G proteins) is coupled to the receptor. G proteins consist of α, β, and γ subunits. When a hormone binds to the receptor, the β and γ subunits dissociate from the α subunit, allowing GDP to be replaced by GTP on the α subunits and causing the β and γ subunits to dissociate from it. The GTP-bound α subunits activate phospholipase C, leading to the release of stored calcium that activates Ca^{2+}-dependent protein kinase (see Figure 11–6).

A. Alpha-adrenergic receptors   Alpha₁ receptors are G$_q$-coupled receptors with three homologous subtypes: α$_{1A}$, α$_{1B}$, and α$_{1C}$. When an agonist binds to the α₁ receptor, the alpha subunit of the guanylyl nucleotide-binding protein, G$_q$, is released and activates phospholipase C. This enzyme catalyzes the conversion of phosphatidylinositol 4,5-bisphosphate to 1,4,5-inositol triphosphate (IP$_3$) and diacylglycerol. IP$_3$ releases calcium ion from intracellular stores to stimulate physiologic responses. Diacylglycerol activates protein kinase C, which in turn phosphorylates a

series of other proteins that initiate or sustain effects stimulated by the release of IP$_3$ and calcium ion (see Figure 11–6). Prazosin and phenoxybenzamine are selective antagonists of the α₁ receptor.

Alpha₁ receptors are postsynaptic (target organ) receptors that mediate diverse effects, particularly arteriolar vasoconstriction. Hepatic α₁ receptors stimulate glycogenolysis and gluconeogenesis. In the eye, α₁ receptors stimulate contraction of the radial muscle of the iris that dilates the pupil. In the skin, α₁ receptors mediate apocrine sweating and pilomotor contraction. In the genitourinary tract, α₁ receptors mediate ejaculation, gravid uterine contraction, and contraction of the bladder sphincter and trigone. In the intestines, α₁ receptors reduce smooth muscle contraction and increase sphincter tone, thereby promoting constipation. Alpha₁ receptors are also found in the splenic capsule and mediate the contraction of the capsule (particularly during hypoxia), which contracts the splenic volume up to 20% and can expel up to 125 mL of packed red blood cells into the circulation within 3 seconds of apnea.

Sympathetic nerves secrete norepinephrine into synaptic junctions where the concentration of norepinephrine is high, thereby stimulating synaptic effector organ α₁ receptors. Epinephrine has a similar effect upon α₁ receptors but is secreted only by the adrenal medulla such that circulating concentrations are usually much lower than concentrations of synaptic norepinephrine. At low plasma concentrations, epinephrine predominantly stimulates β₂-adrenergic receptors (causing vasodilation, discussed later), whereas at higher plasma concentrations, epinephrine stimulates α₁ receptors sufficiently to override vasodilation and cause vasoconstriction.

Alpha₂ receptors are G$_i$ coupled receptors with three highly homologous subtypes: α$_{2A}$, α$_{2B}$, and α$_{2C}$. Agonist binding to the α₂ receptor releases G$_i$ alpha, which inhibits the enzyme adenylyl cyclase and reduces the formation of cAMP (Figure 11–7).

Presynaptic (nerve) α₂ receptors are located near the synapses of sympathetic nerves. Synaptic norepinephrine binds to this receptor, resulting in feedback inhibition of its own release. Alpha₂ receptors are also found in vascular smooth muscle where they mediate vasoconstriction. Alpha₂ receptors in platelets stimulate platelet aggregation and clotting. In the brain, α₂ receptors are found in the locus ceruleus, cerebral cortex, and limbic system. Although their location is postsynaptic, stimulation of the α₂ receptors results in reduced release of norepinephrine. Clonidine is a central α₂ agonist. Yohimbine is a selective antagonist for the α₂ receptor. Phentolamine is an antagonist for both α₁ and α₂ receptors.

In the pancreatic beta cells, activation of α₂ receptors inhibits insulin secretion. This causes glucose intolerance in patients with pheochromocytoma. The gene *Adra2a* encodes the adrenergic α$_{2A}$ receptor. A polymorphism in *Adra2a* increases the number of α$_{2A}$ receptors on beta cells, which reduces insulin secretion and increases the risk of type 2 diabetes mellitus.

B. Beta-adrenergic receptors   Beta-adrenergic receptors are postsynaptic (target organ) cell surface glycoproteins. Agonist binding to β-adrenergic receptors activates adenylyl cyclase via the G$_s$ alpha subunit to increase the production of cAMP, which in

TABLE 11–3 Catecholamine receptors: location and clinical effects following receptor activation.

Catecholamine Receptor	Tissue Location	Clinical Effects Following Receptor Activation
Alpha$_1$	Vascular smooth muscle Liver Eye Skin Prostate Uterus Intestines Spleen capsule	Increases vasoconstriction (increases blood pressure) Increases glycogenolysis and gluconeogenesis Increases ciliary muscle contraction (pupil dilation) Increases pilomotor smooth muscle contraction (erects hairs) Increases contraction and ejaculation Increases gravid uterus contraction Increases sphincter tone and relaxes smooth muscle Contracts spleen volume, expelling blood
Alpha$_2$	Preganglionic nerves Vascular smooth muscle Pancreatic islet cells Blood platelets Adipose cells Brain	Decreases release of neurotransmitter Increases vasoconstriction (increases blood pressure) Decreases release of insulin and glucagon Increases platelet aggregation Decreases lipolysis Decreases norepinephrine release
Beta$_1$	Myocardium Kidney (juxtaglomerular apparatus) Adipose cells Most tissues Nerves	Increases force and rate of contraction Increases secretion of renin Increases lipolysis Increases calorigenesis Increases conduction velocity
Beta$_2$	Vascular smooth muscle Bronchiolar smooth muscle Liver Intestinal smooth muscle Pancreatic islet cells Adipose tissue Muscles Liver and kidney Uterus smooth muscle	Decreases vasoconstriction (increases blood flow) Decreases contraction (bronchial dilation) Increases glycogenolysis and gluconeogenesis Decreases intestinal motility; increases sphincter tone Increases release of insulin and glucagon Increases lipolysis Increases muscle contraction speed and glycogenolysis Increases peripheral conversion of T_4 to T_3 Decreases nongravid uterine contraction (uterine relaxation)
Beta$_3$	Adipose cells Intestinal smooth muscle	Increases lipolysis Increases intestinal motility
Dopamine$_1$	Vascular smooth muscle Renal tubule	Decreases vasoconstriction (vasodilation) Enhances natriuresis
Dopamine$_2$	Sympathetic nerves Pituitary lactotrophs Gastrointestinal tract Brain	Inhibits synaptic release of norepinephrine Inhibits prolactin release Paracrine functions Neurotransmitter

turn converts protein kinase A to its active form. Protein kinase A then phosphorylates a variety of proteins, including enzymes, ion channels, and receptors (see Figure 11–7).

Beta$_1$-adrenergic receptors are located primarily in the heart, kidneys, and adipose tissue. In the heart, β_1 receptors are located in the sinoatrial (SA) and atrioventricular (AV) nodes, atrium, ventricle, and conduction system; activation of these receptors mediates an increase in heart rate and cardiac contraction with an overall increase in cardiac output. Stimulation of β_1 receptors in the ventricles also increases the firing of idioventricular pacemakers such that excessive stimulation predisposes to ventricular ectopy. Triiodothyronine (T_3) increases the total number of cardiac β_1 receptors; this may contribute to the forceful cardiac contraction and tachyarrhythmias that are characteristic of hyperthyroidism. Beta$_1$ receptors in the renal juxtaglomerular apparatus stimulate the release of renin that, in turn, activates the renin-angiotensin system. In fat cells, activation of β_1 receptors increases lipolysis. In nerve cells, activation of β_1 receptors increases nerve conduction velocity. In the brain, β_1 receptors are

found particularly in the nucleus accumbens and the ventral putamen (striatum).

Beta$_2$-adrenergic receptors are particularly activated by epinephrine. They are located mainly in the smooth muscles of arterioles, veins, and bronchi where they mediate relaxation. In the heart, β_2 receptors are located in the same locations as β_1 receptors (discussed earlier) and likewise mediate increases in cardiac rate and contraction. β_2 Receptors are also located in smaller coronary arteries where they mediate dilation of the coronaries, thereby increasing blood flow to cardiac muscle. Vascular β_2 receptors mediate vasodilation of arterioles, particularly in skeletal muscle and the hepatic artery. Stimulation of venous β_2 receptors mediates venous dilation. Stimulation of β_2 receptors also causes relaxation of other smooth muscles, resulting in bronchial dilation and relaxation of the nonpregnant uterus. During exercise or hypoglycemia, epinephrine is secreted and activates hepatic β_2 receptors. This stimulates glycogenolysis and gluconeogenesis for the increased glucose production that is needed for increased cardiac and skeletal muscle activity. Beta$_2$ receptors in skeletal muscle
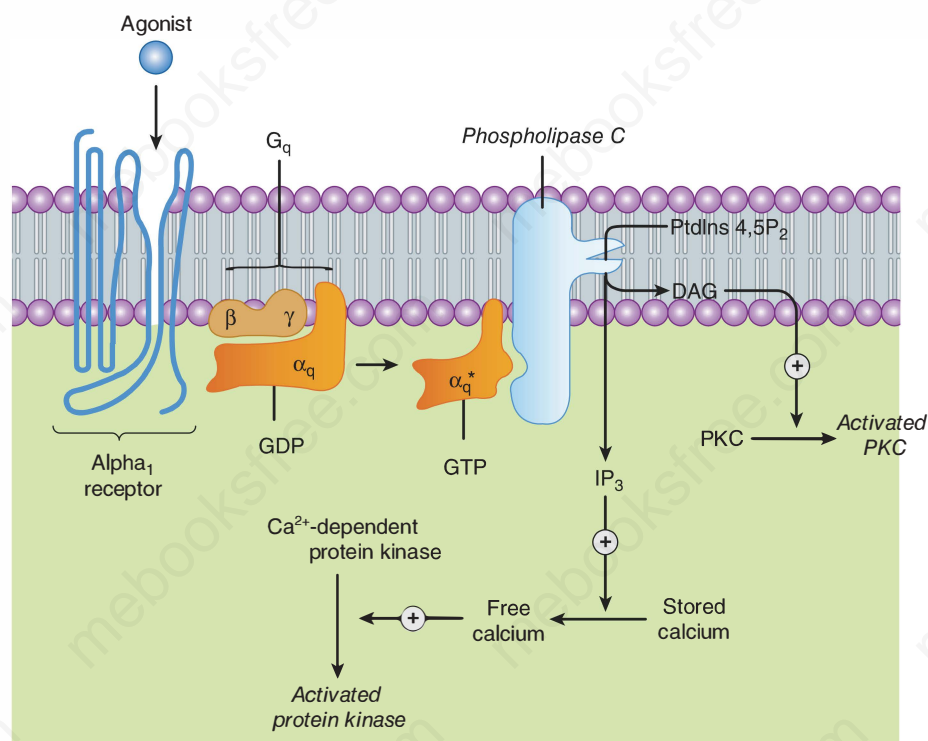
FIGURE 11–6 Activation of α_1-adrenergic responses. Stimulation of α_1 receptors by catecholamines leads to the activation of a G_q-coupling protein. The activated α_q subunit (α_q^*) of this G protein activates phospholipase C, which catalyzes the conversion of PtdIns 4,5-P_2 (phosphatidylinositol 4,5-bisphosphate) to IP_3 (inositol 1,4,5-trisphosphate) and DAG (diacylglycerol). IP_3 stimulates the release of sequestered stores of calcium, leading to an increased concentration of cytoplasmic Ca^{2+} that may then activate Ca^{2+}-dependent protein kinases that phosphorylate their substrates. DAG activates protein kinase C (PKC) (GDP, guanosine diphosphate; GTP, guanosine triphosphate). See Table 11–3 and text for physiologic effects of α_1 receptor activation. (Reproduced with permission from Katzung BJ. *Basic & Clinical Pharmacology*. 12th ed. New York: McGraw-Hill Education; 2012.)

promote increased contraction speed, glycogenolysis, and tremor. The net effect of β_2 agonists (eg, epinephrine) upon liver and skeletal muscle is to increase serum glucose. Beta$_2$-adrenergic receptors are also expressed in various areas of the brain, particularly the right hippocampus, which normally has a higher density of β_2 receptors than the left hippocampus.

Beta$_3$-adrenergic receptors are expressed mainly in adipose tissue but also in the gallbladder, colon, central nervous system, and heart. Activation of this receptor increases energy expenditure, lipolysis, and intestinal motility. β_3 Receptor activation promotes lipolysis in adipose tissue and increased thermogenesis in skeletal muscle. Although β_3-adrenergic receptors are found in the brains of infants, by adulthood their concentration decreases by 100-fold. Pima Indians, who are homozygous for inactivating mutations in the β_3 gene, have an earlier onset of type 2 diabetes.

C. Dopamine receptors   Five different dopamine receptors have been identified (D_1 to D_5), each encoded by a different gene. Dopamine receptors are found throughout the brain (including the anterior pituitary), arteries, proximal renal tubules, and gastrointestinal tract. D_1 and D_2 receptors are the predominant subtypes.

D_1-like family of receptors (D_1 and D_5) couples to G_s that activates adenyl cyclase, thereby increasing intracellular cyclic AMP (cAMP). D_1 receptors mediate vasodilation in the coronary,

renal, cerebrovascular, and mesenteric arteries. D_1 receptors in the proximal renal tubule enhance natriuresis. The D_1 receptor is the most abundant dopamine receptor in the brain. Intravenous therapeutic dopamine at low doses predominantly stimulates D_1 receptors, causing vasodilation. (However, high-dose therapeutic dopamine sufficiently occupies α and β receptors to cause vasoconstriction and increase blood pressure.)

D_2-like family of receptors (D_2, D_3, and D_4) couple to a different G protein, G_i that inhibits adenyl cyclase, thereby reducing intracellular cAMP, opening potassium channels, and decreasing the influx of calcium. D_2 receptors are predominately presynaptic receptors that are located on sympathetic nerves. Activation of D_2 receptors inhibits the synaptic release of norepinephrine and also inhibits the transmission of nerve impulses at sympathetic ganglia. Pituitary lactotroph cells also express D_2 receptors, whose activation inhibits the secretion of prolactin.

Regulation of Sympathoadrenal Activity

The rate of catecholamine secretion largely determines sympathoadrenal activity. However, adrenergic receptors and postreceptor events are sites of fine regulation.

As noted earlier, norepinephrine and epinephrine released during nerve stimulation bind to presynaptic alpha$_2$ receptors and
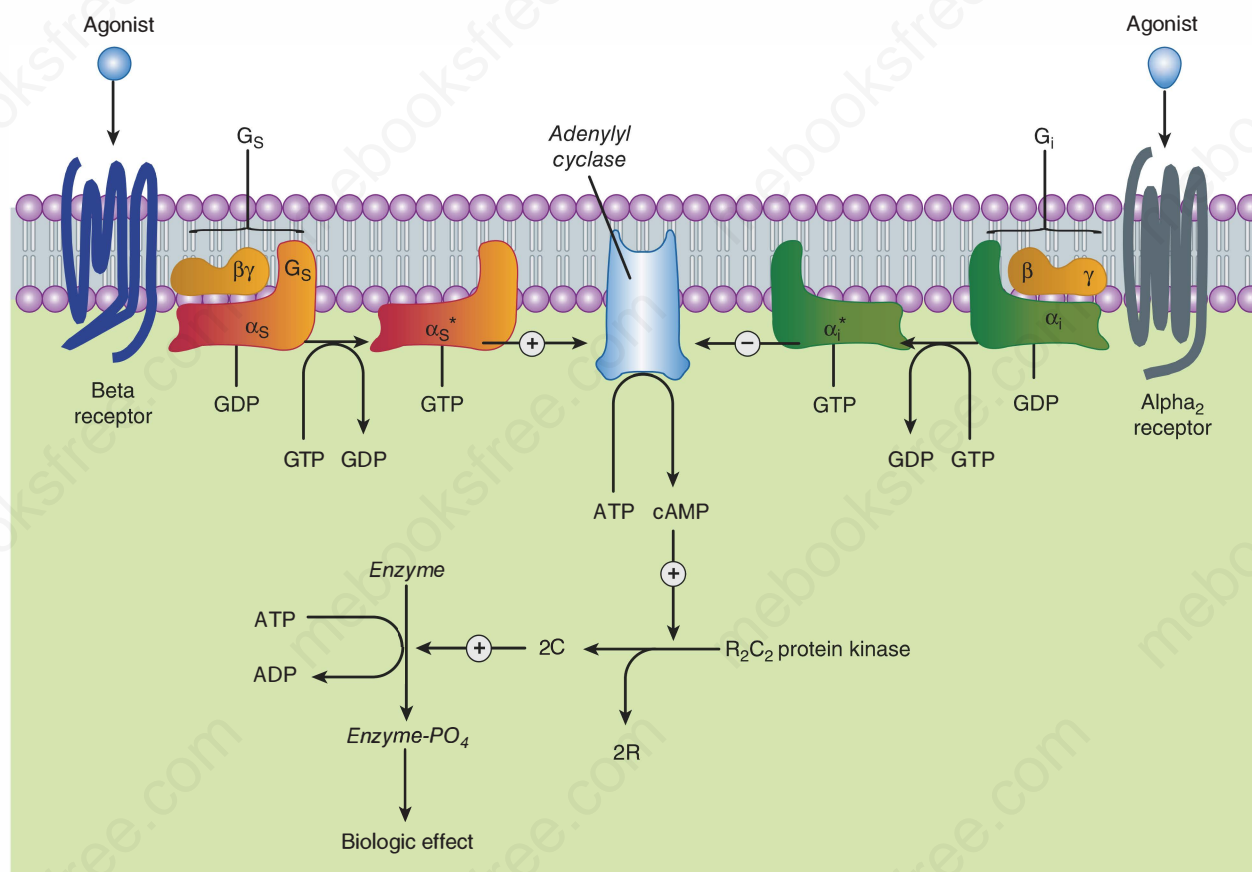
FIGURE 11–7   Activation of β- and α_2-adrenergic responses: activation vs inhibition of adenyl cyclase. Beta receptor ligands activate the stimulatory G protein, G_s, which leads to the dissociation of its $G\alpha_s$ subunit, charged with GTP. The activated $G\alpha_s$ subunit directly activates adenyl cyclase, resulting in increased cAMP that binds to the regulatory subunit (R) of the cAMP-dependent protein kinase, leading to the liberation of active catalytic subunits (C) that phosphorylate specific enzyme substrates and modify their activity. These catalytic units also phosphorylate the cAMP response element–binding protein (CREB), which modifies gene expression. Alpha$_2$ adrenoreceptor ligands inhibit adenyl cyclase by causing dissociation of the inhibitory G protein (G_i) into its subunits; that is, the α_i subunit charged with GTP and a βγ unit. The mechanism by which these subunits inhibit adenyl cyclase is uncertain. See Table 11–3 and text for physiologic effects of β and α_2 receptor activation. (Reproduced with permission from Katzung BJ. *Basic & Clinical Pharmacology*. 12th ed. New York: McGraw-Hill Education; 2012.)

reduce the amount of norepinephrine released. CgA is secreted from neurosecretory vesicles, along with catecholamines. Catestatin, a fragment of CgA, blocks postreceptor cells' cholinergic receptors, thereby inhibiting sympathoadrenal activity.

The binding of a catecholamine agonist to its receptor reduces the number of receptors on the effector cell surface by a process called down regulation. Catecholamine receptor antagonists do not cause down regulation of receptor expression. The mechanisms involved in some of these changes are known. For example, phosphorylation of the β-adrenergic receptor by β-adrenergic receptor kinase results in sequestration of the receptors into membrane vesicles, where they are internalized and degraded. The phosphorylated receptor also has a greater affinity for β-arrestin, another regulatory protein, which prevents its interaction with $G_s\alpha$.

Thyroid hormone increases the number of β receptors in the myocardium. Estrogen increases the number of α-adrenergic receptors in the myometrium and increases the affinity of some vascular α-receptors for norepinephrine.

Actions of Circulating Catecholamines (Tables 11–3 and 11–4)

Norepinephrine activates α_1-adrenergic receptors, increasing the influx of calcium into the target cell. Vascular α_1-adrenergic receptors are found in the heart, papillary dilator muscles, and smooth muscle. Activation of α_1-adrenergic receptors causes hypertension, increased cardiac contraction, relaxation of smooth muscle, and dilation of the pupils. Alpha$_1$-adrenergic activation stimulates sweating from non-thermoregulatory apocrine stress sweat glands, which are located on the palms, axillae, and forehead. Norepinephrine also activates β_1-adrenergic receptors that increase cardiac contraction and rate; stimulation of the heart rate is opposed by simultaneous vagal stimulation. Norepinephrine has less affinity for β_2-adrenergic receptors. However, with higher circulating norepinephrine levels, hypermetabolism and hyperglycemia are noted. Norepinephrine also activates β_3-adrenergic receptors on fat cells, causing lipolysis and an increase in serum levels of free fatty acids.

TABLE 11–4 Plasma levels of catecholamines required to produce hemodynamic and glucose effects.

	Norepinephrine	Epinephrine
Systolic blood pressure	↑ at 2500 pg/mL (15 nmol/L)	↑ at 500 pg/mL (3 nmol/L)
Diastolic blood pressure	↑ at 2500 pg/mL (15 nmol/L)	↓ at 500 pg/mL (3 nmol/L)
Pulse	↓ at 2500 pg/mL (15 nmol/dL)	↑ at 250 pg/mL (1.5 nmol/L)
Plasma glucose	↑ at 2500 pg/mL (15 nmol/L)	↑ at 250-500 pg/mL (1.5-3 nmol/L)

Data from Silverberg AB, et al, *Am J Physiol*. 1978;234:E252; and from Clutter WE, et al, *J Clin Invest*. 1980;66:94.

Epinephrine also stimulates α_1- and β_1-adrenergic receptors, with the same effects noted earlier for norepinephrine. However, epinephrine also activates β_2 receptors, causing vasodilation in skeletal muscles. Epinephrine thus has a variable effect on blood pressure ranging from hypertension to hypotension (rare). Hypoglycemia is a strong stimulus for the adrenal medulla to secrete epinephrine, which increases hepatic glycogenolysis. Epinephrine stimulates lipolysis, resulting in increased serum levels of free fatty acids. Epinephrine also increases the basal metabolic rate. Epinephrine does not cross the blood–brain barrier well. However, high serum levels of epinephrine do stimulate the hypothalamus, resulting in unpleasant sensations ranging from nervousness to an overwhelming sense of impending doom.

Dopamine is an important central neurotransmitter and a precursor to norepinephrine and epinephrine. However, most circulating dopamine is actually derived from the gastrointestinal tract from local synthesis and dietary sources. Dopamine synthesized by the intestines does not function as a neurotransmitter but rather as a paracrine hormone, affecting intestinal motility, blood flow, and sodium transport. Dopamine is also synthesized in the proximal renal tubule where it is a natriuretic hormone. Dopamine synthesis is increased by a high-salt diet. The renal dopamine-paracrine system accounts for the relatively large amounts of dopamine that are normally found in the urine. Circulating dopamine is not normally a significant catecholamine. High serum concentrations of dopamine stimulate vascular D_1 receptors, causing vasodilation and increased renal blood flow. Extremely high serum levels of dopamine are required to sufficiently activate vascular α_1 receptors to cause vasoconstriction.

Physiologic Effects of Catecholamines

A. Cardiovascular effects The release or injection of catecholamines generally increases heart rate and cardiac output and causes peripheral vasoconstriction, leading to an increase in blood pressure. The infusion of catecholamines also leads to a rapid reduction in plasma volume. These events are modulated by reflex mechanisms, so that, as the blood pressure increases, reflex vagal parasympathetic stimulation may slow the heart rate and tend to reduce cardiac output. Although norepinephrine usually has these effects, the effect of epinephrine may vary depending on the smooth muscle tone of the vascular system at the time. For example, in an individual with increased vascular tone, the net effect of small amounts of epinephrine may be to reduce the mean blood pressure through vasodilation, despite an increase in heart rate and cardiac output. In an individual with a reduction in vascular tone, epinephrine increases blood pressure. In addition cardiovascular function is integrated by the central nervous system, so that, under appropriate circumstances, one vascular bed may be dilated while others remain unchanged. The central organization of the sympathetic nervous system is such that its basal regulatory effects are quite discrete—in contrast to periods of stress, when stimulation may be rather generalized and accompanied by release of catecholamines into the circulation.

B. Effects on extravascular smooth muscle Catecholamines have effects on smooth muscle in tissues other than blood vessels. These effects include contraction (α_1) and relaxation (β_2) of uterine myometrium, relaxation of intestinal and bladder smooth muscle (β_2), contraction of the smooth muscle in the bladder (α_1) and intestinal sphincters, relaxation of tracheal smooth muscle (β_2), and pupillary dilation (α_1).

C. Metabolic effects Catecholamines increase oxygen consumption and heat production. This effect is mediated by β_1 receptors. Catecholamines also regulate glucose and fat mobilization from storage depots. Glycogenolysis in heart muscle and liver leads to an increase in available carbohydrate for utilization. Stimulation of adipose tissue leads to lipolysis and the release of free fatty acids and glycerol into the circulation. In humans, these effects are mediated by the β receptor. Both α_1 and β_2 receptors stimulate hepatic glycogenolysis and gluconeogenesis, which causes the release of glucose into the circulation. Both receptors are activated by epinephrine, which is particularly effective in raising hepatic glucose production. Plasma levels of catecholamines required to produce some cardiovascular and metabolic effects in humans are shown in Table 11–4.

Catecholamines have effects on water, sodium, potassium, calcium, and phosphate excretion in the kidney. Stimulation of the β_1 receptor increases the secretion of renin from the renal juxtaglomerular apparatus, thereby activating the renin-angiotensin system. This leads to stimulation of aldosterone secretion.

DISORDERS OF THE ADRENAL MEDULLA AND PARAGANGLIA

EPINEPHRINE AND NOREPINEPHRINE DEFICIENCY

Epinephrine is the major catecholamine secreted by the normal adrenal medulla, and its secretion is unique to the adrenal medulla. Epinephrine is usually deficient in cortisol deficiency of any cause, because high local concentrations of cortisol in the

adrenal medulla are necessary for transcription of the enzyme PNMT, which catalyzes the conversion of norepinephrine to epinephrine in the catecholamine biosynthetic pathway. Primary autoimmune adrenal cortical insufficiency produces epinephrine deficiency. Hypothalamic/pituitary ACTH deficiency also causes cortisol deficiency and consequently epinephrine deficiency.

Autonomic Insufficiency (Table 11–5 and Figure 11–8)

Autonomic insufficiency is also known as dysautonomia. It refers to a general reduction in autonomic activity. Since the adrenal medulla is part of the autonomic nervous system, there is usually an associated deficiency in circulating epinephrine. Autonomic insufficiency most commonly occurs in patients with longstanding type 1 diabetes mellitus. Diabetic autonomic neuropathy may be due to glucose neurotoxicity, but may also be due to autoimmunity. In one study of middle-aged patients with type 1 diabetes mellitus, 52% had complement-fixing antibodies against the sympathetic ganglia and 35% had antibodies against the adrenal medulla. They have minor defects in recovery from insulin-induced hypoglycemia (Figure 11–8), but in patients with diabetes mellitus in whom the glucagon response is also deficient, the additional loss of adrenal medullary response leaves them more susceptible to bouts of severe hypoglycemia. This is the result of a decrease in the warning symptoms as well as an impaired response to the hypoglycemia (see Chapters 17 and 18). Patients with generalized autonomic insufficiency usually have orthostatic hypotension. Familial dysautonomia can manifest as orthostatic tachycardia; norepinephrine transporter deficiency has been determined to be one cause of this condition. The causes of disorders associated with autonomic insufficiency are listed in Table 11–5.

When a normal individual stands, a series of physiologic adjustments occur that maintain blood pressure and ensure adequate circulation to the brain. The initial lowering of the blood pressure stimulates the baroreceptors, which then activate central reflex mechanisms that cause arterial and venous constriction, increase cardiac output, and activate the release of renin and vasopressin. Interruption of afferent, central, or efferent components of this autonomic reflex results in autonomic insufficiency.

Patients with generalized autonomic insufficiency usually have orthostatic hypotension; the reduction in blood pressure on
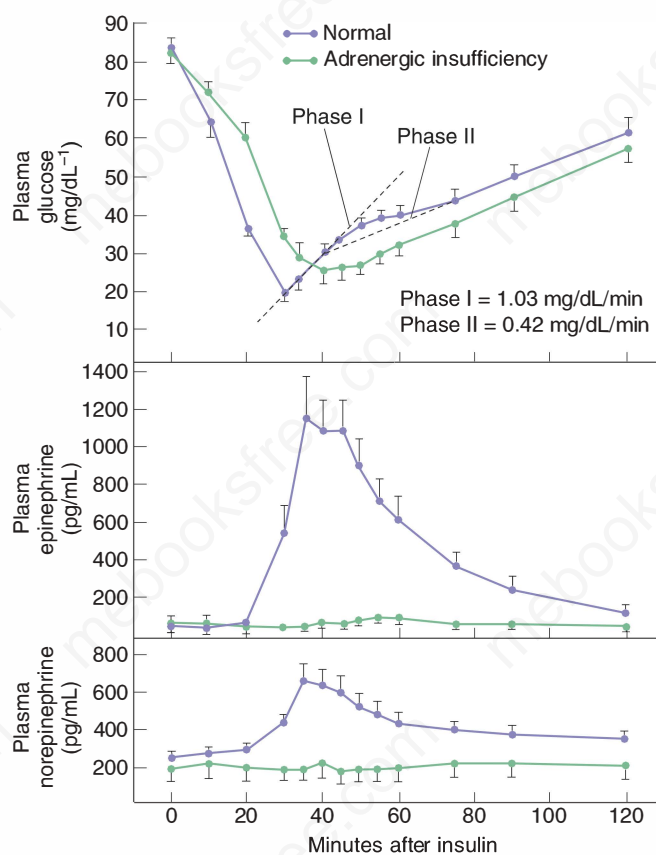
FIGURE 11–8 Plasma glucose, epinephrine, and norepinephrine levels after insulin administration in 14 normal subjects (━●━) and 7 patients with idiopathic orthostatic hypotension (━●━) with low or absent epinephrine responses. Results are expressed as mean ± SEM. (Reproduced with permission from Polinsky RJ et al. The adrenal medullary response to hypoglycemia in patients with orthostatic hypotension. *J Clin Endocrinol Metab*. 1980 Dec;51(6):1401-1406.)

TABLE 11–5 Disorders causing epinephrine or norepinephrine deficiency.

Diabetes mellitus autonomic neuropathy (glucose toxicity or autoimmunity): serum norepinephrine low

Wolfram syndrome: diabetes insipidus, diabetes mellitus, optic atrophy, deafness (DIDMOAD): serum norepinephrine low

Neurological diseases associated with autonomic neuropathy: serum norepinephrine low

Pure autonomic failure: serum norepinephrine and epinephrine low

Primary adrenocortical insufficiency: serum epinephrine low

Congenital adrenal hyperplasia: serum epinephrine low

Allgrove syndrome: achalasia, ACTH resistance, alacrima (AAA syndrome): serum epinephrine low

standing that causes visual disturbances, lightheadedness, and even syncope. Symptoms tend to be worse upon awakening, after meals, in hot weather, and at high altitude. Many patients have supine hypertension, even when not taking hypertensive medication. Affected patients often have a fixed heart rate, sinus tachycardia, diminished sweating, reduced metabolic rate, heat intolerance, nocturia, and erectile dysfunction. Angina pectoris can sometimes occur in the absence of coronary atherosclerosis. Patients frequently have a neurogenic bladder with associated symptoms of frequency, nocturia, hesitancy, reduced urine stream, and recurrent urinary tract infections. Patients may also have gastroparesis and disordered gastrointestinal motility. Sleep apnea is common and death can occur during sleep from reduced respiratory drive.

The diagnosis of autonomic insufficiency can usually be made clinically in patients with typical symptoms and orthostatic blood pressure changes on examination. Additionally, plasma and urine norepinephrine levels are usually very low, sometimes only 10% of average normal in severely affected individuals. Plasma epinephrine levels are often low as well.

The treatment of symptomatic orthostatic hypotension is dependent on maintenance of an adequate blood volume. Physical measures should be employed, such as raising the foot of the bed at night, crossing the legs when sitting, and wearing full-leg fitted support stockings. Adequate hydration is important. **Fludrocortisone** is an effective drug treatment, causing volume expansion. However, fludrocortisone can cause hypokalemia, supine hypertension, and edema. **Midodrine** is an α_1-receptor agonist that must be administered every 8 hours, due to its short half-life of 3 to 4 hours. Midodrine's use is limited by common side effects that include severe supine hypertension, bradycardia, and erythema multiforme. Other adverse effects include pilomotor erection ("goose bumps"), pruritis, and urinary retention. Therefore, midodrine should only be used for patients with very symptomatic orthostatic hypotension who have failed other therapies. Octreotide has been used, either alone or in combination with midodrine, for patients with severe autonomic failure where other treatment measures have failed.

Droxidopa is a vasopressor drug that is converted to norepinephrine by dopa-decarboxylase in peripheral tissues. It is indicated for patients with neurogenic orthostatic hypotension due to autonomic nervous system disorders, including: Parkinson disease, multiple system atrophy, DBH deficiency, pure autonomic failure, and other autonomic neuropathies. Droxidopa induces peripheral arterial and venous vasoconstriction. It is administered orally, starting in doses of 100 mg three times daily and increasing every 1 to 2 days to a maximum of 600 mg three times daily. Droxidopa can cause headache, dizziness, nausea; it also cause supine hypertension that responds to elevating the head and torso during sleep.

PHEOCHROMOCYTOMA AND PARAGANGLIOMA

Both PHEO and non–head-neck paraganglioma (PGL) are tumors of the sympathetic nervous system. Although they are similar tumors, they warrant distinction from each other to emphasize their differences in the following: (1) locations, (2) manifestations, (3) secretion profiles, (4) genetic syndromes, (5) difficulty of surgical resection, and (6) propensity to metastasize.

PHEOs are tumors that arise from the adrenal medulla, whereas non–head-neck PGLs arise from the ganglia of the sympathetic nervous system. PHEOs are more common (85%), and may be bilateral (17%). Patients with adrenal PHEOs are less likely to develop detectable metastases (11%) compared to PGLs (30%).

Prevalence

Among adult patients with hypertension, the prevalence of PHEO/PGL is between 0.2% and 0.6%. In children with hypertension, the prevalence of PHEO/PGL is about 1.7%. Nearly 5% of incidentally discovered adrenal masses are found to be PHEOs; about 9% of patients with such PHEOs are normotensive with normal tumor markers and atypical symptoms, particularly dizziness, or vertigo.

TABLE 11–6 Causes of death in patients with unsuspected pheochromocytomas.

Myocardial infarction
Cerebrovascular accident
Cardiac dysrhythmias
Irreversible shock
Renal failure
Dissecting aortic aneurysm
Acute respiratory distress syndrome

In the general population, PHEO and PGL tumors are diagnosed in about 2 to 8/million population yearly; nearly half of these present as an unexplained death. The National Cancer Registry in Sweden has reported that PHEOs/PGLs are discovered in about 2 patients/million people yearly. Autopsy series suggest a higher incidence, varying from about 250 cases/million to 1300 cases/million in a Mayo Clinic autopsy series. Retrospectively, 61% of PHEO/PGL tumors at autopsy occurred in patients who were known to have had hypertension; about 91% had the nonspecific symptoms associated with secretory PHEO/PGL tumors. In one autopsy study, a large number of patients with these tumors had nonclassic symptoms such as abdominal pain, vomiting, dyspnea, heart failure, hypotension, or sudden death (Table 11–6). Some patients have only intermittent hypertension or are not hypertensive at all, increasing the likelihood of missing the diagnosis. Considering these autopsy data, it is clear that the majority of PHEO/PGL tumors are not diagnosed during life. Tumors occur in both sexes and at any age. They are most commonly diagnosed in the fourth and fifth decades, but tend to present earlier in patients with germline mutations predisposing to PGL.

Pheochromocytomas are chromaffin tumors that arise from the adrenal medulla. They account for 90% of all PHEO/PGL tumors in adults and 70% in children. Adrenal PHEOs are usually unilateral (90%). Unilateral PHEOs occur more frequently in the right (65%) versus the left adrenal (35%). Right-sided PHEOs have been described as producing paroxysmal hypertension more often than sustained hypertension, whereas the opposite was true for tumors arising from the left adrenal. Adrenal PHEOs are bilateral in about 10% of adults and 35% of children. Bilateral PHEOs are particularly common (24% overall) in patients with familial pheochromocytoma syndromes caused by certain germline mutations (discussed later).

Basically, the adrenal medulla is a sympathetic ganglion whose neurons secrete neurotransmitters (epinephrine and norepinephrine) into capillaries instead of synapses. The adrenal medulla is surrounded by the adrenal cortex, with which it shares a portal system that bathes it in high concentrations of cortisol. Cortisol stimulates expression of the gene encoding PNMT, the enzyme that catalyzes the conversion of norepinephrine to epinephrine. Adrenal PHEOs nearly always secrete catecholamines with variable amounts of epinephrine as well as norepinephrine and their metanephrine metabolites. Metastases from adrenal PHEOs usually secrete only norepinephrine and its metabolite normetanephrine; they may rarely secrete epinephrine.

PHEOs are typically encapsulated by a true capsule or a pseudocapsule, the latter being the adrenal capsule. PHEOs are

firm in texture. Hemorrhages that occur within a PHEO can give the tumor a mottled or dark red appearance. A few *black* PHEOs have been described, with the dark pigmentation believed to be due to an accumulation of neuromelanin, a catecholamine metabolite. Larger tumors frequently contain large areas of hemorrhagic necrosis that have undergone cystic degeneration; viable tumor may be found in the cyst wall. Calcifications are often present. PHEOs can invade adjacent organs, and tumors may extend into the adrenal vein and the vena cava, resulting in pulmonary tumor emboli. PHEOs vary tremendously in size, ranging from microscopic to 3600 g. The average PHEO weighs about 100 g and is 4.5 cm in diameter.

Sympathetic PGLs (non head-neck) arise from sympathetic ganglia (Figure 11–9). They account for about 10% of all PHEO/PGL tumors in adults and about 30% in children. About 75% are intra-abdominal where they are often mistaken for adrenal PHEOs. PGLs are sometimes nonsecretory and may be confused with other neuroendocrine tumors. They arise most commonly in the perinephric, periaortic, and bladder regions but often found in the chest (25%), arising in the anterior or posterior mediastinum or the heart. Pelvic PGLs may involve the bladder wall, obstruct the ureters, and metastasize to regional lymph nodes. About 36% to 60% of paragangliomas are functional, secreting norepinephrine and normetanephrine. Functional status is not known to affect survival. Nonfunctional PGLs can often concentrate metaiodobenzylguanidine (MIBG) or secrete normetanephrine and CgA. PGLs of the bladder can cause symptoms associated with micturition; such symptoms occur particularly after voiding large amounts of urine, as upon arising in the morning. Large

perinephric tumors can cause renal artery stenosis and elevations in plasma renin activity. Vaginal tumors may cause dysfunctional vaginal bleeding. PGLs can rarely arise in central nervous system locations, including the sella turcica, petrous ridge, and pineal region. Cauda equina PGLs can cause increased intracranial pressure.

PGLs commonly metastasize (30%-50%) and present with pain or a mass. They tend to metastasize to the liver, lungs, lymph nodes, and bone. PGLs can be locally invasive and may destroy adjacent vertebrae and cause spinal cord and nerve root compression. PGL metastases usually secrete norepinephrine and normetanephrine. They do not secrete epinephrine and therefore do not typically cause significant weight loss, hyperglycemia, anxiety, or tremor.

Head-neck paragangliomas (HNPGL) are nonchromaffin tumors of parasympathetic ganglia. HN-PGLs occur with an incidence of 10 to 33 cases/million persons. Unlike chromaffin sympathetic paragangliomas, only about 5% of these tumors secrete catecholamines. However, about 23% secrete dopamine, and serum dopamine levels can be a useful tumor marker for patients whose dopamine levels are elevated prior to surgical resection. **Carotid body tumors (chemodectomas)** are PGLs that arise near the carotid bifurcation. They usually present as a painless neck mass near the carotid that sometimes has a bruit. Carotid body tumors can injure the vagus or hypoglossal nerves and can metastasize. **Jugulotympanic PGLs** arise in the middle ear and cause tinnitus, hearing loss, and dizziness. The tumor may be seen as a reddish-blue colored pulsating mass behind an intact tympanic membrane (*red drum*) that blanches when pressure is applied with a pneumatic ear speculum (Brown sign). **Glomus jugulare PGLs** arise in the jugular foramen. They often cause tinnitus and hearing loss as well as compression of cranial nerves, resulting in dysphagia. Vagal PGLs are rare, usually presenting as a painless neck mass often associated with hoarseness or dysphagia. Head-neck PGLs can also arise in the larynx, nasal cavity, nasal sinuses, and thyroid gland. Head-neck PGLs themselves are of more interest to otolaryngologists than endocrinologists. But they may be seen along with sympathetic PHEO/PGL in familial PGL syndromes, particularly those with germline SDHD and SDHC mutations (discussed later).

Nonchromaffin HN-PGLs behave differently from sympathetic PHEO/PGLs, although they are embryologically related to chromaffin sympathetic PGLs and may arise concurrently with sympathetic PGLs in familial PGL syndromes. These tumors are less likely to be malignant, although their indolence can be deceiving; recurrence and metastases may not appear for many years. Metastases to local nodes, lungs, and bone can occur. Long-term surveillance for recurrence or metastases is recommended for all patients. These tumors can produce CgA, and preoperative determination of CgA is recommended to determine if it will be a good serum marker for tumor recurrence. About 45% of patients with HN-PGLs carry *SDHx* germline mutations. About 33% harbor SDHD mutations; about 9% have SDHB mutations; about 4% have SDHC mutations. Therefore, gene sequencing is strongly recommended for all patients with HN-PGLs. Individuals harboring such germline mutations are prone to multiple and bilateral
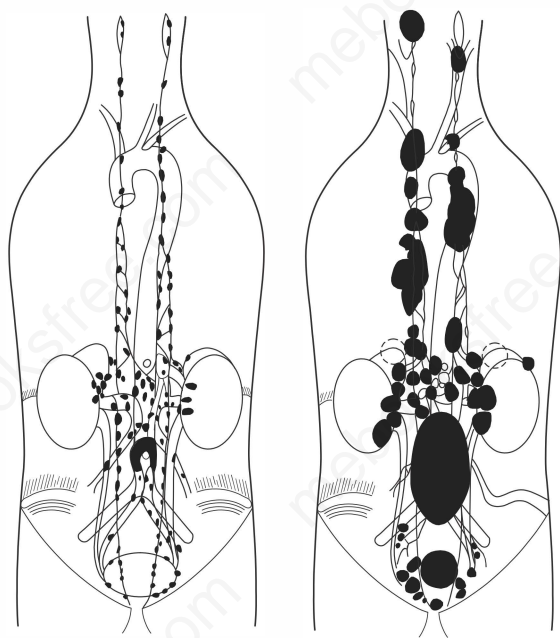
FIGURE 11–9 Left: Anatomic distribution of extra-adrenal chromaffin tissue in the newborn. **Right:** Locations of extra-adrenal PHEOs reported before 1965. (Reproduced with permission from Coupland R. The Natural History of the Chromaffin Cell. *Longmans Green*; 1965.)

HN-PGLs and must also be screened for PGL/PHEO tumors outside the neck (see Genetic Conditions Associated with Pheochromocytomas and Paragangliomas, discussed later).

Neuroblastomas, ganglioneuroblastomas, and ganglioneuromas are sympathetic nervous system tumors that are related to PHEOs and likewise arise from embryonal sympathogonia (see Figure 11–1). Neuroblastoma is the most common malignant disease of infancy and the third most common pediatric cancer, accounting for 15% of childhood cancer deaths.

Ganglioneuroblastomas are composed of a mixture of neuroblasts and more mature gangliocytes; these develop in older children and tend to run a more benign course. Ganglioneuromas are the most benign of these tumors and are composed of gangliocytes and mature stromal cells.

Despite catecholamine secretion, children with neuroblastomas tend to be more symptomatic from their primary tumor or metastases than from catecholamine secretion. Tumors tend to concentrate radiolabeled MIBG, making it a useful imaging and therapeutic agent. Treatment of malignant tumors consists of surgery, chemotherapy, external beam radiation to skeletal metastases, and high-dose ^{131}I-MIBG therapy for patients with MIBG-avid tumors.

Screening for Pheochromocytomas and Paragangliomas

Hypertension is defined in adults as either a systolic blood pressure over 140 mm Hg or a diastolic blood pressure over 90 mm Hg. Hypertension is an extremely common condition, affecting about 20% of all American adults and over 50% of adults over 60 years of age. The incidence of PHEO/PGL is estimated to be less than 0.1% of the entire hypertensive population, but it is higher in certain subgroups whose hypertension is labile or severe. The risk for PHEO/PGL also appears to be higher in patients who have had cyanotic congenital heart disease, possibly due to stimulation of hypoxia-sensitive pathways in the adrenal medulla and paraganglia. Screening for PHEOs should be considered for such patients with severe hypertension and also for hypertensive patients with suspicious symptoms (eg, headaches, palpitations, sweating episodes, or unexplained bouts of abdominal or chest pains; Table 11–7).

Genetic Conditions Associated with Pheochromocytomas and Paragangliomas (Tables 11–8 through 11–12)

PHEOs and **PGLs** usually appear to occur sporadically, with only about 10% to 15% of affected patients readily giving a family history compatible with a familial germline mutation. However, 30% to 40% of patients with PHEO/PGL harbor a germline mutation predisposing to the tumor. This disparity is due to variable penetrance for PHEO/PGL, undiagnosed PHEO/PGL among family members with hypertension and sudden death, and patients' incomplete knowledge of their family's medical history. Mutations in multiple different genes have been demonstrated to predispose to familial PHEO/PGL (Table 11–8). Somatic mutations are

TABLE 11–7 Patients to be screened for pheochromocytoma and paraganglioma.

Hypertension in youth
Hypertensive crisis or shock related to:
    Anesthesia induction
    Drugs: decongestants, glucocorticoids, MAO inhibitors
    Invasive procedures
    Parturition
    Surgery
Hypertensive patients with:
    Symptoms listed in Table 11–14
    Cardiomyopathy
    Cyanotic congenital heart disease
    Erythrocytosis
    Family history of PHEO/PGL or medullary thyroid carcinoma
    Gastrointestinal stromal tumors (GIST)
    Hemangioblastoma
    Hyperglycemia
    Hypertension that is uncontrolled, severe, or markedly labile
    Medullary thyroid carcinoma
    Mucosal neuromas
    Neurofibromatosis and other neurocutaneous syndromes
    Orthostatic hypotension
    Personal history of prior PHEO/PGL
    Pituitary adenoma
    Renal cell carcinoma
    Seizures
    Shock (unexplained)
    Weight loss
Patients harboring germline mutations associated with PHEO or PGL
Radiologic evidence of an adrenal mass
Radiologic evidence of a mass in area of paraganglia

TABLE 11–8 Germline mutations predisposing to pheochromocytoma and paraganglioma.[*]

Gene Mutation
Cluster 1A:
VHL2: von Hippel Lindau Type 2
      VHL Type 2A
      VHL Type 2B
      VHL Type 2C
EGLN2: Prolyl Hydroxylase Domain 1 (PHD1)
EGLN1: Prolyl Hydroxylase Domain 2 (PHD2)
EPAS1-HIF2A: Hypoxia-Inducible Factor-2α (HIF-2α)

Cluster 1B:
SDHA: Succinate Dehydrogenase complex subunit A: PGL5[**]
SDHB: Succinate Dehydrogenase complex subunit B: PGL4[**]
SDHC: Succinate Dehydrogenase complex subunit C: PGL3[**]
SDHD: Succinate Dehydrogenase complex subunit D: PGL1[**]
SDAF2: Succinate Dehydrogenase Assembly Factor 2: PGL2[**]
MDH2: Malate Dehydrogenase 2
FH: Fumarate Hydratase

Cluster 2:
RET: Multiple Endocrine Neoplasia Type 2 (MEN 2)
      MEN 2A
      MEN 2B
NF-1: von Recklinghausen Neurofibromatosis 1
KIF1β: Kinesin Family member 1β
MAX: Myc-Associated factor X
TMEM127: Transmembrane protein 127

[*]See text.

[**]Familial Paraganglioma/Pheochromocytoma (PGL/PHEO) syndrome.

TABLE 11–9 Suggested lifetime surveillance protocol for individuals with a *VHL2* germline mutation.

Patient self-monitors blood pressure regularly and reports hypertension or suspicious symptoms.
Age 10-15 years: Yearly retinal examination and blood pressure measurement; plasma free normetanephrine levels.
Age >15 years: Twice yearly physical examinations with blood pressure measurement; plasma free normetanephrine.
For suspicious symptoms and before major surgical procedures and pregnancy: biochemical screening for pheochromocytoma with plasma free normetanephrine.
Every 2 years: MRI scanning (with intravenous contrast) of the entire brain and spinal cord (VHL types 2A and 2B VHL); MRI scanning of the abdomen.
For abnormal biochemical screening: MRI or CT scan of the abdomen (nonionic contrast) with thin-section adrenal cuts; ^{123}I-MIBG SPECT or ^{18}F-FDA PET scan to confirm the identity of small masses.

TABLE 11–11 Suggested lifetime surveillance protocol for individuals carrying a *RET* proto-oncogene germline mutation.

Patient self-monitors blood pressure regularly and reports hypertension or suspicious symptoms.
Twice yearly: physical examination with blood pressure and neck examination; plasma free metanephrine; serum calcium and albumin (serum PTH if *RET* codon 634 mutation); serum calcitonin (patients with intact thyroid or with medullary thyroid carcinoma).
Prophylactic thyroidectomy: by age 5 years for MEN 2A and by 6 months for codon 918 mutations (MEN 2B).
Yearly thyroid ultrasound for patients with an intact thyroid.
For rising or elevated serum calcitonin levels after thyroidectomy: close surveillance for medullary thyroid carcinoma is required. [18]FDG-PET-CT fusion scans are particularly helpful.
For suspicious symptoms and before major surgical procedures and pregnancy: biochemical screening for pheochromocytoma with plasma free metanephrine.
For abnormal biochemical screening for pheochromocytoma: MRI or CT scan of the abdomen (nonionic contrast) with thin-section adrenal cuts; ^{123}I-MIBG SPECT or ^{18}F-FDA PET scan to confirm the identity of small masses.

common in sporadic PHEO/PGL. An underlying somatic mutation can be detected in 65% to 80% of all cases of PHEO/PGL.

About 20% of patients with an apparently sporadic single PHEO/PGL and no known family history of PHEO/PGL can be shown to harbor a germline mutation, if extensive genetic testing is performed. Such unsuspected germline mutations are more prevalent in PGLs than in PHEOs. For example, about 9% of patients with apparently sporadic, unilateral PHEOs have been found to have von Hippel-Lindau (VHL) mutations. Genetic testing is advisable for all patients with a PHEO/PGL. Genetic testing is strongly recommended for patients with extra-adrenal paragangliomas or multifocal tumors, onset of symptoms before age 45 years, prior history of head-neck PGL, family history of PHEO/PGL or other tumors associated with MEN 2 or VHL. Genetic screening

is also performed for patients with other manifestations of genetic syndromes noted later. Commercial genetic screening is available. Ambry Genetics offers a panel (PGLNext) that screens 12 genes for germline mutations and deletions that predispose to PHEO/PGL: *FH, MAX, MEN 1, NF-1, RET, SDHA, SDHAF2, SDHB, SDHC, SDHD, TMEM127, and VHL. Neurofibromatosis* is usually diagnosed clinically without genetic testing, since the clinical manifestations of NF-1 are usually unmistakable and the *NF-1* gene is very large and costly to sequence.

Known germline mutations that predispose to PHEO/PGL have been categorized into two clusters, according to their reputed common underlying mechanisms of tumorigenesis. Patients who harbor such germline mutations require close surveillance. Suggested lifetime surveillance protocols for patients with the most frequent germline mutations are presented in Tables 11–9 to 11–12.

Cluster 1 germline or somatic mutations involve genes that cause a pseudohypoxia signature that triggers an overabundance of hypoxia inducible factor (HIF), which promotes angiogenesis and cell division. Patients harboring mutations in this cluster might be sensitive to hypoxia. Individuals harboring SDHD germline mutations who live at high altitude have been shown to develop

TABLE 11–10 Suggested lifetime surveillance protocol for individuals with a germline *SDHB, SDHC*, or paternally-inherited *SDHD* mutation.

Patient self-monitors blood pressure regularly and reports hypertension or suspicious symptoms.
Twice yearly: physical examination and blood pressure; plasma-fractionated free metanephrines and serum CgA.
Yearly: scanning with ultrasound of the neck, abdomen, and pelvis.
Every 3 years: MRI of the chest, abdomen, and pelvis to detect early PHEO/PGL and renal cell carcinoma (RCC).
MRI of the head and neck (SDHC, paternal SDHD mutations) to detect head-neck PGL and jugular tympanicum PGL.
For suspicious symptoms and before endoscopies, major surgical procedures, and pregnancy: complete biochemical screening for pheochromocytoma with fractionated plasma metanephrines and serum chromogranin A.
For patients with epigastric discomfort: upper endoscopy, due to incidence of gastrointestinal stromal tumors (GIST).
For patients with headache, amenorrhea, hypogonadism, or clinical suspicion for a pituitary adenoma: biochemical evaluation and MRI for pituitary adenoma.
For abnormal biochemical screening: MRI or CT scan (nonionic contrast) of the chest, abdomen, and pelvis or ^{18}F-FDG-PET with CT fusion scan; to confirm ambiguous results: ^{123}I-MIBG SPECT or ^{18}F-FDA PET scan.

TABLE 11–12 Suggested lifetime surveillance protocol for individuals carrying a *NF-1* gene mutation.

Patient self-monitors blood pressure regularly and reports hypertension or suspicious symptoms.
Twice yearly: physical and neurologic examination and blood pressure, and a careful examination of the skin for development or growth of neurofibromas.
Yearly: CBC with WBC differential; plasma-fractionated free metanephrines; complete eye examination with visual fields.
Before major surgical procedures and pregnancy: plasma-fractionated free metanephrines to screen for pheochromocytoma.
For abnormal biochemical screening: MRI or CT scan of the abdomen (nonionic contrast); ^{123}I-MIBG SPECT or ^{18}F-FDA PET scan.

more PGLs that present earlier in life, compared to other affected members of their kindred living at sea level. However, it is not known whether altitude affects tumorigenesis in kindreds with other germline mutations. Whether other causes of chronic hypoxia (eg, sleep apnea, chronic obstructive pulmonary disease) might have a tumor-promoting effect upon these patients is also unknown. VHL disease is the most common condition associated with Cluster 1A mutations. Mutations in *PHD1, PHD2, and HIF2A* can also predispose to PHEO/PGL. Germline mutations in the gene Cluster 1B mutations encode proteins that are critical for oxidative metabolism in the tricarboxylic acid (TCA) cycle, also known as the citric acid cycle or Krebs cycle. These include mutations in the following: *SDHA, SDHB, SDHC, SDHD, SHAF2, MDH2,* and *FH.*

A. Cluster 1A Mutations

1. **von Hippel-Lindau (*VHL*) 3p25.3** The *VHL* gene is a tumor suppressor gene that encodes two different proteins (pVHL) of 213 and 160 amino acids. Both gene products have a role in the degradation of the hypoxia-inducible factors (HIF-1α and HIF-2α) as follows: the pVHL proteins have an α domain that binds with the protein elongin. The β domain of pVHL is open to bind HIF that has been hydroxylated, a reaction that requires oxygen. This complex then binds ubiquitin that targets HIF for intracellular destruction by proteases. In this manner, well-oxygenated cells destroy HIF. Conversely, cells that are either oxygen-deprived or that lack functional pVHL cause an intracellular accumulation of HIF, a transcription factor that induces the production of vascular endothelial growth factor (VEGF), erythropoietin, erythropoietin receptor, glucose transporter-1, and platelet-derived growth factor-B; these proteins allow an adaptation to hypoxia, but, in excess, they are believed to enhance tumorigenesis.

A germline mutation in the *VHL* tumor suppressor gene has been identified in most families with VHL disease. About 60% of affected families have loss-of-function mutations (30% with truncated pVHL and 30% with large *VHL* gene deletions), resulting in type 1 VHL. About 40% have missense mutations, resulting in an amino acid substitution in pVHL, causing type 2 VHL (discussed later).

In patients with VHL disease, vascular tumors, particularly hemangioblastomas, renal cysts, and renal cell carcinomas, develop when there is a somatic second-hit on the wild-type allele in one cell; this can be caused by various spontaneous mutations (loss of heterozygosity) or promoter hypermethylation of the wild-type VHL allele. In these vascular tumors, a second-hit is usually necessary to cause sufficient accumulation of HIF to promote tumorigenesis. In contrast, PHEOs that develop in patients with type 2 VHL typically have a normal wild-type VHL allele. However, most VHL-associated PHEOs also demonstrate a somatic loss of chromosome 3 (94%) or chromosome 11 (86%).

Although certain VHL kindreds are predisposed to develop PHEOs, it is not the dominant tumor of this syndrome. PHEOs develop only in patients with type 2 VHL (discussed later), and these are different from sporadic PHEOs, in that they are less likely to be malignant, more likely to be bilateral, more likely to be asymptomatic, and and more likely to present at an earlier age. Adrenal medullary hyperplasia is not associated with PHEOs in VHL, in contrast to MEN 2. Although most VHL-associated PHEOs arise in the adrenal, extra-adrenal

sympathetic PGLs sometimes occur and head-neck parasympathetic PGLs have been described. The mean age of presentation for a PHEO/PGL in type 2 VHL is 28 years; the youngest reported patient was 5 years old. In VHL disease, most PHEO/PGLs are relatively small (average 3.6 cm) and are characterized by a thick vascular tumor capsule.

The different types and subtypes of VHL are as follows:

Type 1 VHL—Affected members of families with type 1 VHL do not develop PHEOs. They tend to have loss-of-function *VHL* gene mutations, particularly gene deletions, frameshifts, or truncations.

Type 2 VHL—Affected members of families with type 2 VHL mutations are prone to develop PHEOs. These patients carry VHL missense mutations. Type 2 VHL may be divided into subtypes as follows:
  a. **Type 2A VHL**—PHEOs, hemangioblastomas, *low* risk of renal cell carcinoma.
  b. **Type 2B VHL**—PHEOs, hemangioblastomas, *high* risk of renal cell carcinoma.
  c. **Type 2C VHL**—PHEOs, *no* hemangioblastomas, *no* renal cell carcinomas.

The prevalence of VHL is at least 1/36,000 persons. About 20% of cases are caused by de novo mutations and do not have a family history of VHL. VHL disease carriers are predisposed to multicentric hemangioblastomas in the retina (retinal angiomas), cerebellum, and spinal cord. Renal cysts and renal clear-cell carcinoma commonly develop.

Pancreatic lesions occur in about 70% of patients with VHL. Such lesions are usually discovered incidentally and are usually multiple cysts or cystadenomas. Although such lesions are usually benign, they can cause pancreatitis or jaundice by obstruction of the pancreatic duct or the bile duct, respectively. Pancreatic neuroendocrine tumors (pNETs) occur in up to 17% of cases and are usually multiple. NETs have also arisen outside the pancreas in the duodenum and gallbladder. In VHL disease, NETs tend to occur at a relatively early age of about 35 years. They are usually nonfunctioning. However, they may secrete CgA, which can be a useful tumor marker. Such NETs are usually indolent, particularly those with a diameter under 3 cm and a doubling-time of over 500 days; such NETs are often followed clinically with scanning every 2 years. However, larger and more aggressive NETs are more likely to be malignant and are usually resected. Only 60% are visible with somatostatin receptor scintigraphy.

Other endocrine abnormalities have been described in VHL disease, including hyperparathyroidism and an aggressive pituitary adenoma that secreted both prolactin and growth hormone.

Endolymphatic sac tumors may occur, resulting in hearing loss, vertigo, or ataxia. Adnexal cystadenomas of probable mesonephric origin (APMO) are found in many women with VHL. APMOs may develop in the ovary, broad ligament, vagina, uterine cervix, and vulva. Equivalent epididymal cystadenomas occur in men with VHL.

Genetic testing for VHL disease should be done in children born to a parent with known VHL disease. If VHL disease is suspected, but the family's VHL mutation is not known, the patient must be screened for point mutations by direct *VHL* gene sequencing of the entire coding region and splice junctions.

The clinical diagnosis of VHL disease is made when a patient with a known *VHL* gene mutation develops one tumor

typical of VHL. In patients without a known family history of VHL, a presumptive diagnosis of VHL is made when they develop two or more hemangioblastomas or a hemangioblastoma in association with a PHEO or clear-cell renal carcinoma. Similarly, VHL disease should be suspected in patients with multiple VHL-associated tumors or one VHL-associated tumor that presents at a young age (<50 years for PHEO or hemangioblastoma; <30 years for clear-cell renal carcinoma). Such patients should have *VHL* gene sequencing.

In a French series of 36 patients with PHEOs and VHL disease, PHEOs were the presenting tumor in 53%. PHEOs tended to develop at an early age and were bilateral in 42%; concurrent PGLs were present in 11%. Three of the 36 patients had a malignant PHEO. In 18% of these patients with VHL disease, PHEO was the only known manifestation. Approximately 9% of patients with apparently sporadic, unilateral PHEO have been found to harbor a germline VHL mutation. In certain regions of Europe, the percentage is up to 20%, due to a founder effect for the Tyr98His *black forest* mutation that is more common in kindreds with German ancestry.

PHEOs in VHL disease exclusively produce norepinephrine. Therefore, the metabolite of norepinephrine (normetanephrine) is also produced in these tumors. Plasma free normetanephrine levels are usually elevated when patients with VHL disease develop a PHEO. Therefore, plasma free normetanephrine is the best screening test for patients who harbor a type 2 *VHL* gene missense mutation.

In patients with VHL, a major risk to life is the development of a renal cell carcinoma. When an abdominal computed tomography (CT) scan detects a solid renal lesion, it must be removed. Even simple cystic renal lesions are considered premalignant, and their removal is advisable if renal function can be preserved. If renal cysts are observed, they must be followed every 6 months with dynamic thin-section CT scanning to search for characteristics of malignancy: growth, wall irregularity, or septation. Individuals carrying a *VHL* mutation must have close medical surveillance. A surveillance protocol is recommended in Table 11–9.

2. **Egg Laying Gene Nine (*EGLN*)** These genes encode hypoxia-inducible factor—prolyl hydroxylase (HIF-PH) proteins (ie, prolyl hydroxylase domain [PHD] proteins). PHD proteins are enzymes that hydroxylate hypoxia-inducible factor (HIF), marking it for degradation by the VHL ubiquitination complex. Malfunction of PHD proteins causes an increase in HIF, predisposing to congenital erythrocytosis and formation of PHEO/PGL tumors. PHD proteins exist as three isoforms: PHD1, PHD2, and PHD3, which are encoded by the *EGLN* homolog genes: *EGLN2, EGLN1,* and *EGLN3*, respectively. PHEO/PGL tumors have been described in patients with germline mutations in *EGLN2* and *EGLN1*. An additional mutation of the wild-type *EGLN* allele in an adrenal medulla or paraganglion is required to give rise to a sympathetic PHEO/PGL. In patients harboring *EGLN* germline mutations, available first-degree relatives have tested negative for the mutation, indicating that these may be early post-zygotic mutations. These gene mutations are different from *VHL* gene mutations, although both cause an increase in HIF, and provide an alternative mechanism for PHEO/PGL tumorigenesis.

EGLN2 **19q13.2**: This gene encodes the enzyme prolyl hydroxylase domain 1 (PHD1). Germline mutations in *EGLN2* are rare. The index case was a woman who presented at age 6 with erythrocytosis (borderline elevated serum erythropoietin levels) that required phlebotomies at age 10, hypertension and a PHEO at age 14 that was resected but recurred and metastasized to a thoracic periaortic lymph node. At age 48, she was found to have contralateral PHEOs, an aortocaval PGL, a pelvic PGL, and a bladder PGL.

EGLN1 **1q42.1**: This gene encodes the enzyme prolyl hydroxylase domain 2 (PHD2). Germline mutations in *EGLN1* are rare. The index case was a woman who presented at age 16 with erythrocytosis (borderline elevated serum erythropoietin levels) requiring phlebotomies. At age 39, she was found to have hypertension and a left PHEO and 2 PGLs in the left renal hilum. At age 60, a right PHEO and 3 periaortic PGLs were resected.

3. **Endothelial PAS1 (*EPAS1*) *aka* Hypoxia Inducible Factor 2A (*HIF2A*) 2p21-p16** The *EPAS1*-HIF2A gene encodes the protein eponymously known as endothelial PAS domain-containing 1 (EPAS1) protein or hypoxia inducible factor 2α (HIF-2α). Rare mutations in *EPAS1-HIF2A* cause a syndrome of early PHEO/PGL, somatostatinomas, and erythrocytosis. Mutations causing constitutive activation and prolonged half-life of HIF-2α are typically somatic mutations found in sporadic PHEO/PGL tumor tissue. However, patients with *EPAS1* constitutively activating germline mutations have been described. A familial PHEO/PGL syndrome has not been described with this gene. These appear to be post-zygotic germline mutations that affect only certain related tissues. The index case presented with erythrocytosis in early childhood and high serum erythropoietin levels. At age 35, he was found to have hypertension and a retroperitoneal PGL, followed by multiple additional PGLs with pulmonary and bone metastases. Other patients have been described with multiple PGLs being detected as early as age 14 years. Affected patients are also prone to develop duodenal or pancreatic somatostatinomas and retinal abnormalities with decreased visual acuity.

B. Cluster 1B Mutations: PHEO/PGL syndromes due to mutations involving the tricarboxylic acid cycle (TCA or Krebs cycle) Certain kindreds have a proclivity to develop multicentric head-neck PGLs, sympathetic PGLs, and adrenal PHEOs. They are now also known to be prone to develop renal cell carcinomas and pituitary adenomas. These autosomal dominant syndromes were originally called familial PGL syndromes, although some patients develop adrenal PHEOs. Four of the familial PGL syndromes are caused by mutations in three of the four genes encoding mitochondrial complex II or a gene encoding their assembly into succinate dehydrogenase, which catalyzes the conversion of succinate to fumarate in the mitochondrial Krebs cycle (citric acid cycle) ultimately leading to oxidative phosphorylation and ATP generation. The resultant accumulation of intracellular succinate prevents the degradation of HIF; high intracellular levels of HIF stimulate cell division and angiogenesis, thereby promoting tumorigenesis. The responsible mutations occur in the genes encoding the following: SDH subunit B (*SDHB, PGL 4*), SDH subunit C (*SDHC, PGL 3*), SDH subunit D (*SDHD, PGL 1*), and succinate dehydrogenase complex assembly factor 2 (*SDAF2,* PGL 2). SDHB encodes a catalytic subunit, while *SDHC* and *SDHD* encode membrane-anchoring subunits

involved in electron transport. *SDHC* and *SDHD* (paternal) have an autosomal dominant proclivity to develop PGLs; those with *SDHB* and *SDHD* mutations can also develop adrenal PHEOs. Overall, about 12% of patients with PGLs and PHEOs have been found to harbor one of these germline mutations.

These syndromes are caused by mutations in nuclear genes that encode 3 of the 4 subunits that comprise mitochondrial complex II (SDH) that oxidizes succinate to fumarate (Krebs cycle). The four subunits of the tetrameric SDH consist of a 70-kDa flavoprotein (SDHA), a 30-kDa iron-sulfur protein (SDHB), a 15-kDa subunit of cytochrome b (SDHC), and a 12-kDa subunit of cytochrome b (SDHD). SDHC and SDHD components of cytochrome b are integral mitochondrial membrane subunits anchoring the catalytic subunits SDHA and SDHB that are involved with the electron transport chain, transferring an electron to coenzyme Q (ubiquinone). SDH is essential for aerobic energy production and the TCA cycle. Genetic defects cause mitochondrial dysfunction making the cells functionally hypoxic. This leads to increased secretion of VEGF that is necessary for tumor growth. (*SDHA* germline mutations do not cause PGLs but rather cause Leigh syndrome, a fatal early-onset mitochondrial neurodegenerative disease.)

All patients with PGLs and PHEOs ideally should be tested for germline mutations in *SDH*; such testing is highly recommended for patients with neck PGLs (>15% have germline mutations), other PGLs (particularly multifocal PGLs or PHEOs), a family history of PGL or PHEO, and PGLs arising in patients with Dutch ancestry.

Carney-Stratakis syndrome: Some patients with loss-of-function mutations in *SDHB, SDHC,* or *SDHD* can develop the dyad of PGLs and gastric stromal sarcomas, which is known as **Carney-Stratakis syndrome**. Tumors present relatively early, with an average age of 24 years at diagnosis. Women and men are affected equally. The PGLs that develop in this syndrome are usually functional and develop in the abdomen, rather than head-neck region. In this syndrome, the condition is familial, and affected kindreds do not develop pulmonary chondromas, factors that distinguish it from Carney triad.

1. **Succinate Dehydrogenase complex subunit A (*SDHA*) 5p15**
   SDHA germline mutations are usually lethal. *SDHA* mutations are more typically somatic mutations that have been found in sporadic paragangliomas. This is also known as *PGL 5*.
2. **Succinate Dehydrogenase complex subunit B (*SDHB*) 1p36**
   SDHB germline mutations cause a familial PGL/PHEO syndrome known as *PGL 4*. Affected individuals are prone to develop PGLs all along the parasympathetic and sympathetic chains, from the neck to the pelvis, and can develop adrenal PHEOs as well; they are less prone to glomus tumors of the neck than individuals with *SDHD* mutations. Sympathetic PGLs that arise in patients with *SDHB* germline mutations are much more likely to be metastatic (35%) at the time of diagnosis than tumors seen in patients with *SDHD* germline mutations. In one kindred with a large *SDHB* exon 1 deletion, the phenotypic penetrance of PGL was 35% by age 40. There have been no clear genotype-phenotype correlations with different *SDHB* mutations. However, certain kindreds have a distinctly higher penetrance. Also, some point *SDHB* "mutations" may

be relatively benign genetic polymorphisms rather than pathological mutations. Due to the rarity of these tumors, the precise penetrance for each *SDHB* mutation has not been fully determined. Other malignancies may also be more common in patients with *SDHB* mutations. In one series of 53 patients, renal cell carcinoma was detected in two patients. Pediatric neuroblastoma has been described in a child with an *SDHB* exon deletion.

3. **Succinate Dehydrogenase complex subunit C (*SDHC*) 1q2**
   SDHC encodes a protein of the same name, also known as succinate dehydrogenase cytochrome b560 subunit, mitochondrial. The SDHC protein forms a transmembrane dimer with SDHD that anchors complex II (SDHB and SDHA) to the inner mitochondrial membrane. SDHB germline mutations cause a familial PGL/PHEO syndrome known as *PGL 3*. SDHC germline mutations predispose to parasympathetic HN-PGLs and sympathetic PGLs but not adrenal PHEOs. About 4% of patients with HN-PGLs harbor an *SDHC* germline mutation.
4. **Succinate Dehydrogenase complex subunit D (*SDHD*) 11q23** The resultant protein is also known as succinate dehydrogenase (ubiquinone) cytochrome b small subunit, mitochondrial (CybS). SDHD germline mutations cause a familial PGL/PHEO syndrome known as *PGL 1*. Only patients with paternally inherited *SDHD* gene mutations are predisposed to develop PGLs and PHEOs. Affected individuals are particularly prone to develop parasympathetic HN-PGLs that do not typically secrete catecholamines. About 15% of patients with HN-PGLs harbor this germline mutation. Patients with SDHD mutations are more prone to develop renal cell carcinoma and can also develop Carney-Stratakis syndrome: gastrointestinal stromal cell tumors (GIST); such tumors have an additional acquired somatic SDHD mutation in their normal wild-type gene.

In kindreds with *SDHD* gene mutations, only paternal transmission of the mutated gene causes the susceptibility to PGLs and PHEOs; this phenomenon is known as **maternal genomic imprinting**, meaning that the maternally inherited mutant gene does not cause the syndrome in the mother's offspring. A male who inherits an *SDHD* mutation from his mother does not express the phenotype (no PGLs) but can pass on the gene to his children, who can express the phenotype (PGL/PHEO). A female who inherits an *SDHD* mutation from her father develops PGL/PHEOs, but her children who inherit the *SDHD* mutation are not affected.

Individuals with *SDHD* germline mutations are particularly prone to develop HN-PGLs that arise from parasympathetic ganglia that are embryologically related to sympathetic paraganglia. These tumors do not usually secrete catecholamines and are known as **nonchromaffin** PGLs. Multicentric tumors have been reported in about 74% of PGL patients with *SDHD* mutations. About 50% of patients with seemingly isolated PHEOs and a germline *SDHD* mutation harbor a hidden PGL. Sympathetic PGLs and adrenal PHEOs (8%) also occur in affected individuals.

A founder *SDHD* mutation has been noted in families of Italian descent: *Q109X*. Three *SDHD* founder mutations have been discovered in families of Dutch ancestry: *Leu95Pro*, *Asp92Tyr,* and *Asp92Tyr*. An investigation of 243 family members with a paternally inherited *Asp92Tyr* mutation in SDHD reported the following. The risk of developing a PGL or PHEO was 54% by age 40 years, 68% by 60 years, and 87% by 70 years. Most patients had HN-PGLs, while some had

sympathetic PGLs and 8% had PHEOs. Multiple tumors were found in 65%.

Malignancy appears to be uncommon in patients with SDHD-associated PGLs. However, malignant neck PGLs can be indolent, and metastases to neck nodes, lungs, and bones may not be clinically evident for many years. Therefore, life-long surveillance is necessary (see Table 11–10).

5. **Succinate Dehydrogenase complex Assembly Factor 2 (*SDAF2*) 11q12.2** This gene is also known as *SDH5*. Germ-line mutations in *SDAF2* cause familial PGL syndrome type 2 (PGL 2). The *SDAF2* gene is a member of a family of genes known as mitochondrial respiratory chain complex assembly factors. There appears to be maternal imprinting and only individuals with a paternally inherited mutated gene develop both PGLs and PHEOs. The most common mutation has been Gly78Arg.

6. **Malate Dehydrogenase 2 (*MDH2*) 7cen-q22** The *MDH2* gene encodes an enzyme of the TCA cycle. The index case for this germline mutation was a 55-year-old man who presented with multiple malignant PGLs.

7. **Fumarate Hydratase (*FH*) 1q42.1** The *FH* gene encodes another enzyme in the TCA cycle. Germline *FH* mutations predispose to familial leiomyomatosis and renal cell carcinoma (FLRCC). Affected individuals develop cutaneous and uterine leiomyomatosis type 2 papillary renal cell carcinoma, and PGLs. About 40% of PGLs have presented with metastatic disease.

Cluster 2 germline or somatic mutations cause an overexpression of RAS/MAPK and P13K-AKT-mTOR signaling pathways. Cluster 2 mutations include the following: (1) *RET* proto-oncogene activating mutations (Multiple Endocrine Neoplasia Type 2A, MEN 2A and MEN 2B); (2) von Recklinghausen Neurofibromatosis Type 1 (*NF-1*); (3) *KIF1β*; (4) *MAX*; and (5) *TMEM127*.

1. ***RET* proto-oncogene 10q11.2** *RET* is an abbreviation for the term "rearranged during transfection." Germline activating mutations in the *RET* proto-oncogene give rise to **multiple endocrine neoplasia type 2 (MEN 2A and MEN 2B)**.

The first description of PHEO was in a patient with MEN 2. In 1882, Felix Fränkel described PHEOs in an 18-year-old woman with bilateral adrenal tumors; genetic testing of four living relatives has revealed a germline mutation in the *RET* proto-oncogene.

The prevalence of MEN 2 has been estimated at 1 in 30,000. MEN 2 is an autosomal dominant disorder that causes a predisposition to medullary thyroid carcinoma, PHEO, duodenal or pancreatic somatostatinomas (rare), and other abnormalities described later. MEN 2 is caused by an activating mutation in the *RET* proto-oncogene on chromosome 10q11.2. *RET* consists of 21 exons that encode a transmembrane receptor tyrosine kinase that is expressed in neural crest tissues. The constitutive activation of *RET* causes hyperplasia in affected tissues, including the adrenal medulla. Additional mutation(s) are believed to be required for a PHEO to develop. The somatic loss of a tumor suppressor gene on chromosome 1p appears to be necessary for PHEO formation. Additionally, reduced expression of *NF-1* has been reported in a minority of PHEOs arising in patients with MEN 2 mutations.

MEN 2 kindreds can be grouped into two distinct subtypes: MEN 2A (90%) and MEN 2B (10%). Patients with MEN 2A have various single amino acid missense mutations affecting the extracellular *RET* domain that cause *RET* homodimerization and constitutive activation of its tyrosine kinase. Patients with MEN 2B have a particular missense mutation (codon 918, exon 16) affecting the intracellular domain at the *RET* tyrosine kinase's catalytic site, causing constitutive activation. In either subtype, PHEOs usually develop in the adrenals (96%); extra-adrenal paragangliomas are rare (4%). About 42% of patients with PHEO in MEN 2 have hypertension, usually paroxysmal. Overall, only about 53% exhibit symptoms of PHEO. Each specific type of mutation in the *RET* codon determines each kindred's idiosyncrasies, such as the age at onset and the aggressiveness of medullary thyroid carcinoma.

A limited number of *RET* gene exons have been found to harbor mutations that are capable of causing constitutive activation of the tyrosine kinase. *RET* exons 10, 11, 13, 14, 15, and 16 are usually involved. Therefore, routine genetic screening searches for mutations only in these exons. If no mutation is found in these exons, the remaining 15 exons can be sequenced in a research laboratory. When an affected kindred's *RET* mutation is already known, all first-degree relatives in the kindred should be screened for that specific mutation.

PHEOs arise in the adrenals, and extra-adrenal PGLs are uncommon in MEN 2. In patients with MEN 2, only about 4% of PHEOs are found to be metastatic, possibly because of earlier detection. When PHEOs develop in patients with MEN 2, they are bilateral in about two-thirds of cases; however, with close surveillance of affected kindreds, PHEOs are discovered earlier and are more likely to be unilateral. After a unilateral adrenalectomy for PHEO in a patient with MEN 2, a contralateral PHEO develops in about 50% of patients an average of 12 years after the first adrenalectomy. In patients with MEN 2, adrenal PHEOs produce norepinephrine and epinephrine (with its metabolite metanephrine). When screening for small PHEOs, plasma catecholamine concentrations may be normal, but plasma metanephrine concentrations are usually elevated, making plasma free metanephrines the screening test of choice for these patients. However, some PHEOs are detected only by a 24-hour urine determination for fractionated metanephrines, catecholamines, and creatinine.

MEN 2A (Sipple syndrome)—Patients with this genetic condition develop medullary thyroid carcinoma (95%-100%), hyperparathyroidism due to multiglandular hyperplasia (20%), and PHEO (50%; range 6%-100% depending on the kindred) or adrenal medullary hyperplasia. Patients with MEN 2A also have an increased incidence of cutaneous lichen amyloidosis and Hirschsprung disease.

Individuals belonging to an MEN 2A kindred should have genetic testing for *RET* proto-oncogene mutations by 5 years of age to determine if they carry the genetic mutation that will require prophylactic thyroidectomy and close surveillance for PHEO and hyperparathyroidism.

More than 85% of the mutations in MEN 2A families affect codon 634 in exon 11 of the *RET* proto-oncogene. PHEOs tend to present in middle age, often without hypertension. The exceptions are patients who harbor a *RET* C634R (arginine) germline mutation, in whom 59% develop bilateral PHEO by age 40 and who therefore require more intense surveillance for PHEO at an earlier age. Individuals with mutations in RET codon 630 also have a high incidence of PHEO. PHEOs also occur in most other kindreds with MEN 2A. Screening for PHEO should commence with the routine blood pressure measurements that are performed during examinations in childhood for medullary thyroid

carcinoma and follow-up of hypothyroidism following thyroidectomy. At about age 15 years, affected individuals should commence yearly screening for PHEO with plasma free metanephrine determinations.

Certain *RET* mutations (codons 609, 768, V804M, and 891) rarely produce PHEOs. Less intense screening for PHEOs is required in patients with these mutations.

MEN 2B—Over 90% of patients with MEN 2B have a single amino acid substitution of methionine to threonine on codon 918 in exon 16 that affects the intracellular domain of tyrosine kinase. About 50% of such mutations are familial while the rest arise de novo as post-zygotic mutations and are sporadic. Patients with this genetic condition are prone to develop mucosal neuromas, PHEO, adrenal medullary hyperplasia, and aggressive medullary thyroid carcinoma. About 50% of affected patients develop all three manifestations of the syndrome. Mucosal neuromas tend to develop first and occur in most patients. They appear as small bumps on the tongue, lips and buccal mucosa, eyelids, cornea, and conjunctivae. The lips and eyelids may become diffusely thickened. Intestinal ganglioneuromatosis occurs and alters intestinal motility, causing diarrhea or constipation and occasionally megacolon. Affected individuals have a marfanoid habitus with associated spinal scoliosis or kyphosis, pectus excavatum, a high-arched foot and talipes equinovarus (club foot) deformities. In patients with MEN 2B, medullary thyroid carcinoma tends to be aggressive and occurs at an earlier age than in patients with MEN 2A. Members of kindred with MEN 2B should immediately have genetic testing for their family's *RET* proto-oncogene mutation. If an individual is found to carry the family's *RET* mutation, early prophylactic thyroidectomy is advisable. In kindreds with MEN 2B, infants are screened for the mutation at birth. For affected infants, prophylactic thyroidectomy is performed by 6 months of age.

All individuals carrying a *RET* proto-oncogene mutation require close lifetime medical surveillance (see Table 11–11). See Chapter 22, Multiple Endocrine Neoplasia.

2. Neurofibromatosis Type 1 (*NF-1*) 17q11.2 PHEOs are ultimately diagnosed in 0.1% to 5.7% of patients with von Recklinghausen neurofibromatosis type 1 (NF-1). Most of these PHEOs are not diagnosed during life, since the autopsy incidence of PHEO in NF-1 patients is 3.3% to 13%. PHEOs that develop in patients with NF-1 are similar to sporadic PHEOs: 84% have solitary adrenal tumors, 10% have bilateral adrenal tumors, 6% have extra-adrenal PGLs, and 12% have metastases or local invasion.

PHEOs are present in 20% to 50% of NF-1 patients with hypertension, and all NF-1 patients with hypertension must be screened for PHEO. It is prudent to screen all patients with NF-1 for PHEO yearly, with interval testing if hypertension or symptoms develop that are suggestive of PHEO (headache, perspiration, palpitations). Similarly, all patients with NF-1 should be screened for PHEO before major surgical procedures and pregnancy.

In patients with NF-1, PHEOs can present anytime during life, from infancy to old age, with a mean age of 42 years at diagnosis. These PHEOs can grow to large size. Although patients with PHEOs may develop hypertension, many patients are surprisingly asymptomatic despite increased catecholamine secretion. Patients with NF-1 are prone to develop vascular anomalies such as coarctation of the aorta and renal artery dysplasia, which can produce hypertension and mimic a PHEO.

von Recklinghausen disease is caused by a mutation in the NF-1 tumor suppressor gene mapped to chromosome 17q11.2. The *NF-1* gene encodes a 2818 amino acid protein called neurofibromin, which inhibits *Ras* oncogene activity; loss of neurofibromin leads to *Ras* activation and tumor formation. It is a common autosomal dominant genetic disorder (although 50% of cases seem sporadic), with an approximate incidence of 300 to 400 cases per million population. Although genetic testing for NF-1 is available, the diagnosis is usually made clinically during childhood or adolescence.

Patients with *NF-1* mutations can present in childhood with optic gliomas that impair vision or in adolescence with plexiform neurofibromas. Patients develop visible subcutaneous neurofibromas and schwannomas of cranial and vertebral nerve roots. Skeletal abnormalities are common. Hypothalamic hamartomas may occur and cause precocious puberty. Iris hamartomas may occur and are known as Lisch nodules. NF-1 patients have an increased risk of developing other tumors, especially malignant peripheral nerve sheath tumors and leukemia (particularly juvenile chronic myelogenous leukemia). Affected patients can also develop pancreatic somatostatinomas. Patients may also have freckles in their axillae and skin folds. Multiple cutaneous pigmented café au lait spots develop and grow in size and number with age; most patients ultimately develop more than six spots (smooth bordered) measuring more than 1.5 cm in diameter.

Individuals carrying an *NF-1* mutation must have lifetime medical surveillance (see Table 11–12).

3. Kinesin Factor 1B (*KIF1B*) 1p36 *KIF1B* is a gene that is believed to synthesize a proapoptotic factor for sympathetic cell precursors. The gene is located on 1p36, where gene deletions have been noted in neural crest tumors. Rare mutations in *KIF1B* have been associated with PHEOs, PGLs, and neuroblastomas, as well as other tumors that are not of neural crest origin. These tumors are transcriptionally related to *NF-1* and *RET* tumors.

4. Myc-Associated factor X (*MAX*) 14q23 Germline mutations in the *MAX* gene predispose to PHEO/PGL. These rare mutations appear to arise from differential paternal transmission in this syndrome. About two-thirds of cases have been found to have multiple tumors or bilateral adrenal tumors that are synchronous or metachronous; such tumors display loss of heterozygosity in the wild-type *MAX* allele. About 10% to 37% of such tumors have metastasized by the time they are discovered.

5. Transmembrane protein 127 (*TMEM127*) 2q11.2 This gene encodes a tumor suppressor protein that controls a signaling pathway (mTORC1) that promotes cell proliferation and survival. Individuals with rare *TMEM127* germline mutations are prone to develop renal cell carcinomas and PHEOs/PGLs; such tumors have been found to have a second somatic mutation in *TMEM127*, resulting in loss of heterozygosity. Affected individuals usually present between ages 42-54 years, older than patients in some other familial PHEO/PGL syndromes. They usually develop adrenal PHEOs that are benign. About one-third of cases have bilateral adrenal PHEOs when discovered. However, secretory PGLs and HN-PGLs have been reported. Only about 25% of patients with *TMEM127* germline mutations give a family history of PHEO, so 75% appear to be "sporadic." *TMEM127* germline mutations are found in about 2% of all patients with PHEO/PGL. No other tumors have been definitely associated with germline *TMEM127* mutations.

Other genetic syndromes associated with PHEO or PGL:

1. **Carney-Stratakis syndrome** This familial syndrome is autosomal dominant and almost always associated with a detectable germline mutation in *SDHB*, *SDHC*, or *SDHD*. Both sexes are affected equally. Affected patients develop PHEO/PGL and gastrointestinal stromal cell tumors (GIST); such tumors have an additional acquired somatic SDHx mutation in their normal wild-type gene. The syndrome is also known as GIST-PGL dyad. This syndrome is *not* associated with pulmonary chondromas.

2. **Carney triad** Multicentric paragangliomas occur in patients with Carney triad. About 150 cases have been reported worldwide. About 10% of affected patients have been found to harbor germline mutations in *SDHA, SDHB,* or *SDHC* genes. However, for the remaining patients, the precise underlying genetic defect is unknown. This condition is not usually familial. Affected individuals are usually young women who develop malignant but indolent GISTs (leiomyosarcomas), pulmonary chondromas, and PHEO/PGLs. Only one-fifth of reported cases have had all three tumors, while most patients have had two of the three tumors, usually gastric sarcomas and pulmonary chondromas. Other tumors can develop in this syndrome, including ganglioneuromas, neuroblastomas, adrenocortical adenomas (12%), and esophageal leiomyomas. Women account for 86% of the cases. Tumors tend to develop early in life, with an average age of 21 years at diagnosis. This tumor syndrome is distinct from Carney-Stratakis syndrome (discussed earlier).

3. **Beckwith-Wiedemann syndrome** PHEOs have been reported in patients with Beckwith-Wiedemann syndrome and may be bilateral. Affected individuals also have other abnormalities, particularly neonatal hypoglycemia, omphalocele, umbilical hernia, macroglossia, and gigantism, and they are prone to develop malignancies.

Other genetic tumor syndromes that may possibly predispose to PHEO/PGL:

1. *ATM* **(Ataxia-Telangiectasia Mutated gene)** Homozygotes have ataxia-telangiectasia, immunodeficiency with respiratory infections, and an increased risk for breast cancer and lymphoid malignancies. At least one patient with an *ATM* germline mutation has developed a PHEO/PGL (author's experience).

2. *SETD2* **(Su[var], Enhancer of zeste, Trithorax-Domain containing 2)** Germline loss-of-function mutations in the *SETD2* gene predispose to renal cell carcinoma and childhood brain gliomas. At least one patient with an *SETD2* germline mutation has developed a PHEO/PGL (author's experience).

Somatic Mutations in PHEO and PGL

Somatic mutations have been found in nearly 50% of PHEO/PGLs. Somatic activating *HRAS* gene mutations have been found only in sporadic PHEO/PGL tumor tissue. The many different germline mutations that cause familial syndromes also occur as somatic mutations in sporadic PHEO/PGL tumors: (1) Activating somatic mutations in *RET and HIF2*; (2) Loss-of-function somatic mutations are predominantly seen for *NF-1*, and somatic mutations in *VHL, SDHB, SDHD,* and *MAX* are less common.

Somatic mutations have also been described for *EPAS1*. Somatic mutations in other genes have been reported in only a few PHEO/PGLs: *MEN 1, EGLN1, EGLN2, MDH2,* and *IDH1*. Other somatic mutations have been described in sporadic PHEO/PGL, but their role in tumorigenesis is uncertain: *TP53, ATRX,* and *BRAF*.

Physiology of PHEO and PGL

PHEOs are tumors that arise from the adrenal medulla, whereas non–head-neck PGLs arise from sympathetic nerve ganglia. Although some PHEOs/PGLs do not secrete catecholamines, most synthesize catecholamines at increased rates that may be up to 27 times the synthetic rate of the normal adrenal medulla (Figure 11–10). The persistent hypersecretion of catecholamines by most PHEOs/PGLs is probably due to lack of feedback inhibition at the level of tyrosine hydroxylase. Catecholamines are produced in quantities that greatly exceed the vesicular storage capacity and accumulate in the cytoplasm. Such catecholamines and their metabolites (metanephrine and normetanephrine) diffuse out of the PHEO/PGL tumor cells into the circulation.

In contrast to the normal adrenal medulla, most adrenal PHEOs produce more norepinephrine than epinephrine, while many produce both hormones. Some PHEOs produce epinephrine almost exclusively. Most PGLs produce mostly norepinephrine with some dopamine and no epinephrine. PGLs that produce large amounts of dopamine tend to be metastatic. Some PGLs secrete no catecholamines or metanephrines.

Serum levels of catecholamines correlate only modestly with tumor size. On average, large tumors secrete more catecholamines than smaller tumors. However, large tumors are much more variable in their secretion of tumor markers. This may be due to the fact that larger tumors tend to develop hemorrhagic necrosis and cysts that are not functional. Additionally, about 20% of large, malignant PGLs are less differentiated and are nonsecretory.

Severe hypertensive episodes occur in most patients with PHEOs and secretory PGLs. Exocytosis of catecholamines from the PHEO can play a role in such paroxysms, but most PHEOs have minor sympathetic innervation. Instead, hypertensive crises are often caused by spontaneous hemorrhages within the tumor or by pressure on the tumor causing the release of blood from venous sinusoids that are rich in catecholamines. Thus, catecholamines can be released by physical stimuli such as bending or twisting or by micturition in patients with bladder PGLs. Of course, surgical manipulation of such tumors releases catecholamines and can cause life-threatening hypertensive crises.

Chronically high circulating levels of catecholamines may cause normal sympathetic axons to become saturated with catecholamines due to active catecholamine neuronal uptake. This may account for the paroxysms of hypertension that are triggered by pain, emotional upset, intubation, anesthesia, or surgical skin incisions. Adrenergic catecholamine saturation may also explain the elevations in plasma and urine catecholamines that can occur for 10 days or longer after a successful surgical resection of a PHEO/PGL.

	pg/mL	
	Norepinephrine (nl <500 pg/mL)	**Epinephrine** (nl <250 pg/mL)
Superior vena cava	5500	100
Inferior vena cava	5700	530
Above right adrenal vein	6000	1200
Left adrenal vein	6400	47,000
Tumor		
Left renal vein	32,000	2300
Iliac vein	5000	150
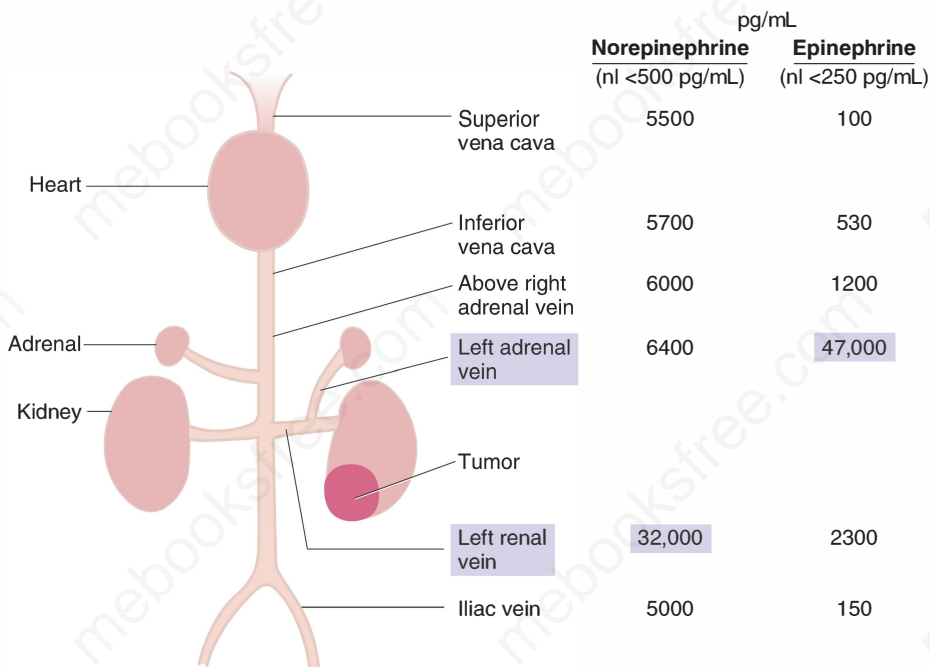
FIGURE 11–10 Plasma norepinephrine and epinephrine levels in venous samples of blood from a patient with a secretory left perinephric PGL. Note that the level of epinephrine is very high in the left adrenal vein sample, distinguishing it from the drainage of the PGL that secretes norepinephrine into the left renal vein. Note the relatively normal peripheral levels of epinephrine (seen in the iliac vein or the superior vena cava). Such venous sampling is no longer required for diagnosis and localization of PHEO or PGL.

Secretion of Other Peptides by PHEOs and PGLs (Table 11–13)

Although PHEOs/PGLs secrete mainly catecholamines and their metabolites, they also secrete many other peptide hormones, many of which contribute to a patient's clinical symptoms. Secretion of parathyroid hormone–related peptide (PTHrP) can cause hypercalcemia. Ectopic ACTH production can cause Cushing syndrome. Erythropoietin secretion can cause erythrocytosis. Leukocytosis is frequently seen in patients with PHEO, probably caused by cytokine release from the tumor. Interleukin-6 secretion can cause fevers and acute respiratory distress syndrome (ARDS).

Chromogranin A: Chromogranins are acidic single-chain glycoproteins that are found in neurosecretory granules. They have been categorized into three classes: chromogranins A (CgA), CgB (secretogranin I), and CgC (secretogranin II). In humans, the gene for CgA is on chromosome 14 and encodes a molecule with 431 to 445 amino acids. CgA molecules aggregate with low pH or high calcium concentrations and promote the formation of secretory vesicles and the concentration of hormones within these vesicles.

CgA also acts as a prohormone. It is cleaved by endopeptidases into smaller peptides. CgA is the prohormone for the amino terminal fragments vasostatin I (CgA 1-76) and vasostatin II (CgA 1-115), which appear to inhibit vasoconstriction. CgA is also the prohormone for catestatin (CgA 352-372), which blocks acetylcholine receptors and thus inhibits adrenosympathetic activity.

Most PHEO/PGL tumors secrete CgA, and serum CgA levels may be assayed as a tumor marker for PHEO. However, CgA is

TABLE 11–13 Other proteins and peptides that may be secreted by pheochromocytomas and paragangliomas, in addition to catecholamines and metanephrines.[a]

Adrenocorticotropic hormone
Adrenomedullin
Atrial natriuretic factor
Beta-endorphin
Calbindin
Calcitonin
Cholecystokinin
Chromogranin A
Cytokines
Enkephalins
Erythropoietin
Galanin
Gonadotropin-releasing hormone
Growth hormone
Interleukin-6 (IL-6)
Motilin
Neuron-specific enolase
Neuropeptide Y
Neurotensin
Parathyroid hormone–related peptide
Peptide histidine-isoleucine
Renin
Serotonin
Somatostatin
Substance P
Tumor necrosis factor alpha (TNFα, cachexin)
Vasoactive intestinal polypeptide

[a]Pheochromocytomas and paragangliomas variably secrete different peptides. Not all peptides have been documented to produce clinical manifestations (see text).

not unique to the adrenal medulla. It is produced in neuroendocrine cells outside the adrenal medulla that secrete peptide hormones. CgA is found in the pituitary gland, parathyroid gland, central nervous system, and pancreatic islet cells.

Neuropeptide Y (NPY): Many PHEOs secrete significant amounts of NPY. The molecule is a 36 amino acid peptide (see primary sequence, discussed later) that is found in neurosecretory granules and is secreted along with catecholamines. It is a very potent nonadrenergic vasoconstrictor that contributes to hypertension in some patients with PHEO.

Tyr-Pro-Ser-Lys-Pro-Asp-Asn-Pro-Gly-Glu-Asp-AlaPro-Ala-Glu-Asp-Met-Ala-Arg-Tyr-Tyr-Ser-Ala-LeuArg-His-Tyr-Ile-Asn-Leu-Ile-Thr-Arg-Gln-Arg-Tyr-NH$_2$

NPY acts on G protein–coupled receptors, which belong to the pancreatic polypeptide family of cell-surface receptors called peptide YY (PYY). Six subtypes of these receptors have been identified (Y$_1$R-Y$_6$R). NPY has particular affinity for the Y$_1$-R, Y$_2$-R, Y$_3$-R, and Y$_5$-R receptors. Some PYY receptors are found in vascular smooth muscle and mediate NPY-induced vasoconstriction. Other PYY receptors are found on the pheochromocytes themselves and mediate NPY's autocrine effects: NPY appears to inhibit catecholamine synthesis via Y$_3$R, a receptor which blocks L-type Ca^{2+} channels. It inhibits catecholamine release via Y$_2$R, a receptor that blocks N-type Ca^{2+} channels.

NPY also has paracrine effects, stimulating endothelial cell proliferation and angiogenesis that may foster tumor growth. Patients with essential hypertension have normal levels of NPY. However, in a series of eight patients with PHEO, NPY levels were elevated 2- to 465-fold above the normal reference range. In another series, 59% of adrenal PHEOs were found to secrete NPY during surgical resection; high serum levels of NPY were observed to correlate with vascular resistance, independent of norepinephrine. NPY appears to contribute to hypertension and left ventricular hypertrophy in most patients with PHEO. In contrast, few PGLs secrete NPY.

Adrenomedullin (AM): It was originally isolated from a PHEO, thus its misleading name. Although AM is produced by the adrenal medulla, it is also produced in the adrenal cortex (zona glomerulosa). It is produced by so many other tissues that the adrenal gland is actually a minor source for circulating AM.

AM is a 52 amino acid peptide cleavage product of a prohormone, preproadrenomedullin, encoded on chromosome 11. AM has some homology with *calcitonin* gene–related peptide (CGRP), and exerts its effects as a ligand for CGRP1 receptors and specific AM receptors. Ligand binding to the G protein–coupled receptors causes adenylate cyclase activation and an elevation of cellular cAMP levels.

AM is a pluripotent hormone. In the adrenal, AM appears to suppress aldosterone secretion. It is also secreted by the heart, lung, kidney, and brain as well as vascular endothelium, where it causes vasodilation. AM is also a natriuretic peptide and is secreted by the heart in congestive heart failure. AM is expressed in PHEOs and a variety of other tumors, where it appears to stimulate tumor growth, reduce apoptosis, and suppress the immune response to the tumor.

Neuron-specific enolase (NSE): It is a neuroendocrine glycolytic enzyme. Serum levels of NSE have been reported to be normal in patients with benign PHEO but are elevated in about half of patients with malignant PHEOs. Therefore, an elevated serum level of NSE may increase the likelihood that a given PHEO is malignant.

Manifestations of PHEO and PGL (Table 11–14)

More than one-third of PHEOs and sympathetic PGLs cause death prior to diagnosis. Death usually results from a fatal cardiac dysrhythmia, myocardial infarction, stroke, cardiomyopathy, ARDS, or multisystem crisis. Adult patients with a secretory PHEO/PGL usually have paroxysmal symptoms, which may last minutes or hours; symptoms usually begin abruptly and subside slowly. The particular constellation of symptoms varies considerably among patients. One cause for the differences in symptoms is the variable production of epinephrine and norepinephrine by these tumors. PHEOs that produce epinephrine tend to cause paroxysmal β-adrenergic manifestations, particularly anxiety, tremor, diaphoresis, tachycardia, palpitations, and hyperglycemia. Epinephrine and cytokine secretion can cause pulmonary edema and ARDS. PGLs do not secrete epinephrine but most secrete norepinephrine that causes hypertension. PGLs are more likely to metastasize.

Manifestations and their approximate incidence include hypertension (90%), headaches (80%), diaphoresis (70%), and palpitations or tachycardia (60%). Other common symptoms include episodic anxiety (60%), tremor (40%), abdominal or chest pain (35%), pallor (30%), and nausea or vomiting (30%). Hyperglycemia occurs in about 30% but is usually asymptomatic; diabetic ketoacidosis has been reported but is very rare. Patients may also experience fever (28%), fatigue (25%), flushing (18%), and dyspnea (15%). Change in bowel habits occurs frequently with either constipation (13%) or diarrhea (6%). Visual changes occur in 12% with either transient blurring or field loss during attacks; metastases to the orbit or skull base may directly impinge on the optic nerve (see Table 11–14).

Triggers for paroxysms: Episodic paroxysms may not recur for months or may recur many times daily. Each patient tends to have a different pattern of symptoms, with the frequency or severity of episodes usually increasing over time. Attacks can occur spontaneously or may occur with bladder catheterization, anesthesia, or surgery. Acute attacks may also be triggered by eating foods containing tyramine: aged cheeses, meats, fish, beer, wine, chocolate, or bananas. Hypertensive crises can also be triggered by certain drugs: radiocontrast media (especially ionic type), MAO inhibitors, tricyclic antidepressants, sympathomimetics, decongestants, glucagon, chemotherapy, glucocorticoids (prednisone, dexamethasone), ACTH, opiates, metoclopramide, nicotine, and cocaine. Phenothiazines have been reported to cause shock and pulmonary edema in patients with PHEO.

Paroxysms can be induced by seemingly benign activities such as bending, rolling over in bed, exertion, abdominal palpation, or micturition (with bladder paragangliomas). There is an amazing

TABLE 11–14 Clinical manifestations of pheochromocytoma and paraganglioma.

Blood pressure	Hypertension: severe or mild, paroxysmal or sustained; orthostasis; hypotension/shock; normotension
Vasospasm	Cyanosis, Raynaud syndrome, gangrene; severe radial artery vasospasm with thready pulse; falsely low blood pressure by radial artery transducer
Multisystem crisis	Severe hypertension/hypotension, fever, encephalopathy, ARDS, renal failure, hepatic failure, death
Cardiovascular	Palpitations, dysrhythmias, chest pain, acute coronary syndrome, cardiomyopathy, heart failure, cardiac paragangliomas
Gastrointestinal	Abdominal pain, nausea, vomiting, weight loss, intestinal ischemia; pancreatitis, cholecystitis, jaundice; rupture of abdominal aneurysm; constipation, toxic megacolon
Metabolic	Hyperglycemia/diabetes; lactic acidosis; fevers
Neurologic	Headache, paresthesias, numbness, dizziness, CVA, TIA, hemiplegia, hemianopsia, seizures, hemorrhagic stroke; skull metastases may impinge on brain structures, optic nerve, or other cranial nerves; spinal metastases may impinge on cord or nerve roots
Pulmonary	Dyspnea; hypoxia from ARDS
Psychiatric	Anxiety (attacks or constant); depression; chronic fatigue; psychosis
Renal	Renal insufficiency, nephrotic syndrome, malignant nephrosclerosis; large tumors often involve the kidneys and renal vessels
Skin	Apocrine sweating during paroxysms, drenching sweats as attack subsides; eczema; mottled cyanosis during paroxysm
Ectopic hormone production	ACTH (Cushing syndrome); VIP (Verner-Morrison syndrome); PTHrP (hypercalcemia)
Children	More commonly have sustained hypertension, diaphoresis, visual changes, polyuria/polydipsia, seizures, edematous or cyanotic hands; more commonly harbor germline mutations, multiple tumors, and paragangliomas
Women	More symptomatic than men: more frequent headache, weight loss, numbness, dizziness, tremor, anxiety, and fatigue
Pregnancy	Hypertension mimicking eclampsia; hypertensive multisystem crisis during vaginal delivery; postpartum shock or fever; high mortality
General laboratory	Leukocytosis, erythrocytosis, eosinophilia
Associated tumors	Renal cell carcinoma, hemangioblastoma, gastric sarcoma, pulmonary chondroma, pituitary adenoma, papillary thyroid cancer

inter-individual variability in the manifestations of PHEO/PGLs. Most patients have dramatic symptoms, but other patients with incidentally discovered secretory PHEOs are completely asymptomatic. Patients who develop PHEOs as part of MEN 2 or VHL disease are especially prone to be normotensive and asymptomatic.

Blood pressure: Hypertensive crisis is the quintessential manifestation of PHEO/PGL. Blood pressure that exceeds 200/120 mm Hg is an immediate threat to life, being associated with encephalopathy or stroke, cardiac ischemia or infarction, pulmonary edema, aortic dissection, rhabdomyolysis, lactic acidosis, and renal insufficiency.

Definition of hypertension: In adults, hypertension is considered to be present when blood pressure exceeds 140 mm Hg systolic or 90 mm Hg diastolic. In children, blood pressure increases with age, such that maximal normal ranges are age dependent; hypertension is considered to be present when blood pressure measurements exceed the following: younger than 6 months, 110/60 mm Hg; 3 years, 112/80 mm Hg; 5 years, 115/84 mm Hg; 10 years, 130/90 mm Hg; and 15 years, 138/90 mm Hg.

Presentation of hypertension: Hypertension is present in 90% of patients in whom a PHEO/PGL is diagnosed. Hypertension occurs primarily from the secretion of norepinephrine. Epinephrine secretion variably increases blood pressure. Also, PHEO/PGLs often secrete neuropeptide Y that is a potent vasoconstrictor.

Some tumors may grow so large that they impinge upon the renal artery, causing increased renin production and secondary renovascular hypertension. Blood pressure patterns vary among patients with PHEOs. Adults most commonly have sustained but variable hypertension, with severe hypertension during symptomatic episodes. Paroxysms of severe hypertension occur in about 50% of adults and in about 8% of children with PHEO. Other patients may be completely normotensive, may be normotensive between paroxysms, or may have stable sustained hypertension.

Hypertension can be mild or severe and resistant to treatment. Severe hypertension may be noted during induction of anesthesia for unrelated surgeries. Although hypertension usually accompanies paroxysmal symptoms and may be elicited by the earlier activities, this is not always the case.

Orthostasis: Patients with sustained hypertension frequently exhibit orthostatic changes in blood pressure, often with orthostatic tachycardia. Patients may complain of orthostatic faintness. Blood pressure may drop, even to hypotensive levels, after the patient arises from a supine position and stands for 3 minutes; such orthostasis, especially when accompanied by a rise in heart rate, is suggestive of PHEO. Epinephrine secretion from a PHEO may cause episodic hypotension and even syncope.

Hypotension and shock: Although hypertension is usually the key symptom in PHEO/PGL patients, hypotension and shock can occur. Some PHEOs produce purely epinephrine that can

produce mild hypertension from alpha stimulation but can also produce hypotension from predominantly β-stimulated vasodilation. After an especially intense and prolonged attack of hypertension, shock may ultimately occur. This may be due to loss of vascular tone, low plasma volume, arrhythmias, or cardiac damage. Spontaneous necrosis within a PHEO can lead to severe hypotension when norepinephrine levels suddenly drop. Similarly, surgical resection of a PHEO/PGL often precipitates sudden and severe intraoperative hypotension, particularly in the presence of α-blockade and other antihypertensives. Cardiogenic shock can occur as a result of cardiac ischemia or infarction as well as catecholamine cardiomyopathy (see Cardiac Manifestations, discussed later).

Peripheral vasospasm and gangrene: Vasoconstriction is responsible for the pallor and mottled cyanosis that can occur with paroxysms of hypertension. Raynaud phenomenon can occur. Peripheral pulses may become thready or even nonpalpable during paroxysms. Catheters inserted into the radial artery and connected to continuous blood pressure monitoring transducers can give misleadingly low-pressure readings during paroxysms of vasospasm, a condition known as **pseudoshock**. Prolonged severe peripheral vasospasm has rarely caused gangrene of the skin, fingers, or toes. Reflex vasodilation usually follows an attack and can cause facial flushing.

Multisystem crisis: Massive release of catecholamines and cytokines from PHEO/PGL tumors can occur spontaneously during tumor necrosis or can be triggered by any of the factors noted earlier. This can cause multisystem crisis that can be the presenting manifestation of PHEO/PGL, with hypertension or hypotension, high fever, encephalopathy, cardiomyopathy with heart failure, ARDS, renal failure, hepatic injury, and death. Multisystem crisis resembles septic shock, so the diagnosis of PHEO may be missed entirely.

Cardiac manifestations: The heart is affected in most patients with PHEO/PGL, either directly from excessive catecholamines or indirectly from hypertension. Palpitations are one of the most frequent complaints.

Dysrhythmias: Tachycardia or dysrhythmias occur in about 60% of patients with PHEO/PGL. Patients may note palpitations, described variably as episodes of excessive heart pounding at rest or a fluttering sensation, during which they may become lightheaded. Supraventricular tachycardia is common, particularly in patients with an epinephrine-secreting adrenal PHEO. The heart rate will often increase when standing. There can be an initial tachycardia during a paroxysm, followed by a reflex bradycardia. Atrial fibrillation may occur. Other dysrhythmias described with PHEO/PGL include nodal tachycardia, torsades de pointes, sick sinus syndrome, and Wolff-Parkinson-White (WPW) syndrome. Atrioventricular dissociation can occur and right bundle branch block has been reported. Ventricular tachycardia or ventricular fibrillation can occur suddenly and is a common cause of death.

Acute coronary syndrome (ACS): During a paroxysm, severe coronary vasospasm can cause myocardial ischemia or infarction, even in the absence of coronary atherosclerosis. This typically occurs simultaneously with increased myocardial oxygen demand caused by catecholamine-induced increases in heart rate and contractility. Patients may experience crushing chest pain or pressure with referred pain, usually to the jaw or the left shoulder or arm. Acute heart failure may occur along with severe hypotension or shock.

It can be difficult to distinguish ACS due to coronary stenosis/thrombosis from ACS due to a PHEO. Changes in the electrocardiogram and serum levels of troponin and creatine kinase-MB (CK-MB) are the same for both conditions. Patients with coronary disease can have tachycardia, cardiomyopathy, hypertension, dysrhythmias, diaphoresis, anxiety, and increased plasma catecholamine levels. However, patients with PHEO usually do not have critical coronary artery stenosis on coronary angiogram and usually have severe hypertension, headache, and other paroxysmal symptoms.

Cardiomyopathy: Left ventricular hypertrophy and hypertensive cardiomyopathy can occur in patients with chronic hypertension from a PHEO/PGL. High levels of catecholamines can also directly cause myocarditis and a dilated cardiomyopathy. This is known as **catecholamine cardiomyopathy**. Takotsubo cardiomyopathy is the term used to describe asymmetric cardiac contractility with apical and midventricular akinesis or dyskinesis with hyperkinesis at the base. Other patterns of cardiac hypokinesis can occur. Most patients with catecholamine cardiomyopathy develop pulmonary edema and die. However, full recovery from cardiomyopathy may occur after treatment and surgical resection of a PHEO/PGL. In some patients, myocardial scarring and fibrosis lead to irreversible cardiomyopathy and heart failure. Fatal Takotsubo cardiomyopathy has been reported after dobutamine stress echocardiography.

There can be multiple causes of cardiomyopathy in patients with PHEO/PGL: hypertensive, ischemic, and catecholamine. The underlying pathophysiology of catecholamine cardiomyopathy appears to involve high levels of intracellular calcium in cardiac myocytes. Postmortem histopathology typically reveals intracardiac hemorrhages, edema, and concentrations of lymphocytes and leukocytes in cardiac muscle with areas of myocardial fibrosis.

Cardiac PGLs: These are rare tumors that arise at the base of the heart near the trunks of the aortic and pulmonary arteries and may be asymptomatic or cause hypertension and palpitations. They typically involve the left atrium or interatrial groove and often protrude into the atrium, resembling an atrial myxoma. Cardiac PGLs often have an invasive intramural component, making tumor resection difficult.

Diaphoresis and fever: During a paroxysm, sweating usually occurs initially from apocrine glands, affecting the palms, axillae, head, and shoulders. Reflex thermoregulatory eccrine sweating occurs later in an attack, dissipating the heat that was acquired during the prolonged vasoconstriction that occurred during the paroxysm. This can cause drenching sweats, usually as a paroxysm subsides. Patients with PHEO commonly develop fevers that may be mild or severe, even as high as 41°C. Up to 70% of patients have unexplained low-grade elevations in temperature of 0.5°C or more. Such fevers have been attributed to the secretion of interleukin-6 and respond to nonsteroidal anti-inflammatory drug therapy.

Gastrointestinal manifestations: Many patients with PHEO/PGL exhibit gastrointestinal symptoms such as abdominal pain, nausea, vomiting, and constipation. Most patients lose some weight, even if appetite is preserved. More severe weight loss (>10% of basal weight) occurs in about 15% of patients overall and in 41% of those with sustained and prolonged hypertension. Abdominal pain may be due to splanchnic vasoconstriction (*intestinal angina*) and prolonged vasospasm can cause ischemic enterocolitis. Vasospasm can rarely cause ischemic gangrene of the bowel. Pain may also be caused by the growth of a large intra-abdominal tumor. Intestinal motility disorders are also common. Catecholamines relax gastrointestinal smooth muscle while increasing contraction of the pyloric and ileocecal sphincters. Constipation is common, and abdominal distention and even toxic megacolon can occur. Other abdominal emergencies can be seen with PHEO, including the rupture of an aortic aneurysm, acute cholecystitis, and acute pancreatitis.

Neurologic manifestations: Headache is a common manifestation during an acute paroxysm. Patients frequently complain of paresthesia, numbness, or dizziness. Affected patients have an increase risk of experiencing a cerebrovascular accident (CVA) or transient ischemic attack (TIA). Hemiplegia can occur, sometimes with homonymous hemianopsia, and may be transient or permanent. Although most CVAs are ischemic in origin, hemorrhagic stroke can also occur during hypertensive paroxysms. Rupture of the internal carotid artery into the cavernous sinus can occur, causing a third nerve palsy. Some patients develop confusion or even psychosis during paroxysms. Paresthesias can occur.

Pulmonary manifestations: Patients may complain of dyspnea during a paroxysm. Patients with catecholamine cardiomyopathy may also present with dyspnea. Some patients develop ARDS, which can develop acutely or over several days. This is a life-threatening condition but may be self-limited during an attack. PHEO/PGL can produce ARDS that may be mistaken for pneumonia, pulmonary edema, pulmonary emboli, or congestive heart failure. It is hypothesized that ARDS may be caused by interleukin-6 produced by the tumor. Congestive heart failure may also cause pulmonary edema and can be distinguished from ARDS by echocardiogram.

Renal manifestations: Some degree of renal insufficiency is common in patients with PHEO. Hypertensive nephrosclerosis occurs in patients with a long history of severe hypertension. Nephrotic syndrome may occur with significant proteinuria, possibly due to secretion of interleukin-6 by the tumor. Malignant nephrosclerosis can occur with severe hypertension damaging renal arterioles, resulting in rapidly progressive renal failure. Large PHEOs and perirenal PGLs can impinge upon the renal artery, causing increased renin production and a Goldblatt kidney, resulting in renovascular hypertension that is additive to norepinephrine-induced hypertension. The tumor may directly invade the renal vein and extend into the inferior vena cava, causing pulmonary emboli and an increased risk of lower extremity deep vein thrombosis. Severe hypertensive paroxysms can cause muscle ischemic damage and rhabdomyolysis with release of myoglobin that causes myoglobinuric renal failure. Acute tubular necrosis can occur after a severe hypotensive episode.

Ectopic hormone production: PHEOs can rarely produce ACTH that stimulates the adrenal cortex to produce excessive cortisol, resulting in Cushing syndrome. Tumors can also produce VIP that can cause watery diarrhea, hypokalemia, and achlorhydria, the WHDA (Verner-Morrison) syndrome. Some tumors produce PTHrP that causes hypercalcemia. Although PHEOs can secrete renin directly, elevated plasma renin levels are usually derived from the juxtaglomerular apparatus, whose β_1 receptors are stimulated by both epinephrine and norepinephrine.

Manifestations in children: The symptoms of PHEO/PGL in children are different from those in adults. Over 80% of children exhibit hypertension that is usually sustained and less frequently (10%) paroxysmal. Children are more prone to diaphoresis and visual changes. Children are more likely to have paroxysms of nausea, vomiting, and headache, which often occur after exertion. They are also prone to weight loss, polydipsia, polyuria, and convulsions. Affected children may also exhibit reddish-blue mottled skin along with edematous and cyanotic-appearing hands, a symptom rarely seen in adults. Children are more likely to have multiple tumors and extra-adrenal PGLs. In one series, 39% of affected children had bilateral adrenal PHEOs, an adrenal PHEO plus a PGL, or multiple PGLs; a single PGL occurred in an additional 14% of children. Affected children often have genetic conditions associated with PHEOs/PGLs. Thus, they may harbor the other tumors associated with these conditions. All affected children should be considered for genetic testing for germline mutations or deletions predisposing to PHEO/PGL (see Table 11–8).

Manifestations in women: There are sex differences in PHEO/PGL symptomatology, with women tending to be more symptomatic than men. Women report significantly more headache (80% vs 52%), weight change (88% vs 43%), numbness (57% vs 24%), dizziness (83% vs 39%), tremor (64% vs 33%), anxiety (85% vs 50%), and changes in energy level (89% vs 64%).

Manifestations in pregnancy: A PHEO/PGL during pregnancy can cause sustained hypertension or paroxysmal hypertension that is typically mistaken for eclampsia. Hypertensive paroxysms tend to occur more frequently as the uterus enlarges, triggered by direct pressure upon the tumor by shifts in position or movement of the fetus. Hypertensive crisis typically occurs at the time of vaginal delivery and is commonly associated with cardiac arrhythmia, ARDS, and death. Postpartum women may develop shock or fever that can mimic uterine rupture, amniotic fluid embolus, or infection (puerperal sepsis). A tumor that is unrecognized carries a grave prognosis, with a reported 40% maternal mortality and a 56% fetal mortality. If the diagnosis of PHEO is made before delivery, the maternal mortality rate drops to about 10% (see Pregnancy and Pheochromocytoma, discussed later).

Manifestations of malignancy: Metastases occur in about 11% of adrenal PHEOs and 30% to 50% of sympathetic PGLs. Since histopathology cannot distinguish whether a given PHEO is malignant, the term malignant is dependent upon whether metastases are detectable at presentation (50%) or months to years later (50%). Metastases are usually (80%) functional and can cause recurrent hypertension and symptoms many months or years after an operation that had been thought to be curative (see Metastatic Pheochromocytoma and Paraganglioma, discussed later).

Normotension despite high plasma levels of norepinephrine: Interestingly, about 14% of patients with PHEO/PGL have no hypertension despite having chronically elevated serum norepinephrine levels. This phenomenon has been variably called desensitization, tolerance, or tachyphylaxis.

Patients can be genetically prone to adrenergic desensitization. Adrenergic desensitization is caused by adrenergic receptors undergoing sequestration, down-regulation, or phosphorylation. Adrenergic desensitization appears to be one cause of the cardiovascular collapse that can occur abruptly following the surgical removal of a PHEO/PGL in some patients.

Desensitization does not account for all patients who are normotensive in the face of elevated serum levels of norepinephrine, because some such patients can still have hypertensive responses to norepinephrine. Secretion of epinephrine can have a hypotensive effect and may account, at least in part, for this phenomenon. Some patients are homozygous for certain polymorphisms of β_2-adrenergic receptors that allow continued β_2-adrenergic-mediated vasodilation, thus counteracting the pressor effects of circulating epinephrine and norepinephrine caused by stimulation of vascular α_1-adrenergic receptors.

Cosecretion of dopamine may reduce blood pressure through a central nervous system action. Also, cosecretion of dopamine may directly dilate mesenteric and renal vessels and thus modulate the effects of norepinephrine.

Biochemical Testing for Pheochromocytoma

No single test is absolutely sensitive and specific for PHEO. Plasma-fractionated free metanephrines or urinary 24-hour fractionated metanephrines have a sensitivity of about 97%. Sensitivities of other tests are somewhat lower: urinary norepinephrine 93%, plasma norepinephrine 92%, urinary VMA 90%, plasma epinephrine 67%, urinary epinephrine 64%, and plasma dopamine 63%. However, some malignant tumors secrete only dopamine and no catecholamines and no metanephrines at all. Also, the determination of plasma catecholamines-metanephrine ratios can be of value in discriminating false-positive from true-positive results (discussed later). Therefore, assays for plasma catecholamines, dopamine, and serum chromogranin A are often warranted.

PHEO/PGLs are deadly tumors and missing the diagnosis can be disastrous, so screening tests must be very sensitive. The secretion of catecholamines can be paroxysmal, with low secretion rates between paroxysms. In contrast, the secretion of metanephrine or normetanephrine metabolites is relatively high and constant. Tumors secrete metanephrines in their unconjugated (free) form. Thus, plasma-fractionated free metanephrines is the single most sensitive screening tests for these tumors. Urinary 24-hour fractionated metanephrines has a similar sensitivity but the collection is less convenient. Some rare tumors (usually malignant) have a defect in the conversion of dopamine to norepinephrine by DBH, such that serum dopamine levels are very high while catecholamines are normal or mildly elevated and metanephrines are totally normal. Additionally, some paragangliomas secrete no catecholamines or metanephrines but do secrete CgA.

The establishment of normal reference ranges is problematic for catecholamines and metanephrines, since levels vary with sex, age, and medical conditions:

1. **Sex** Women have lower plasma epinephrine and metanephrine levels than men; their urinary excretion of catecholamines and metanephrines is also lower. Women also have lower 24-hour urine metanephrine and normetanephrine than do men. However, such sex differences disappear when normalized to creatinine excretion.
2. **Age** Children, especially boys, normally have somewhat higher levels of plasma epinephrine and metanephrine than do adults. For boys ages 5 to 17 years, the upper limit of normal (ULN) for plasma free metanephrine is less than 57 pg/mL (0.52 nmol/L) and the ULN for plasma free normetanephrine is 97 pg/mL (0.53 nmol/L). For girls ages 5 to 17 years, the ULN for plasma free metanephrine is 0.37 nmol/L and the ULN for plasma free normetanephrine is 77 pg/mL (0.42 pmol/L). Conversely, children's average 24-hour urine epinephrine and norepinephrine excretion rates are lower than those of adults and increase through childhood as weight increases. Therefore, using ratios of catecholamines to creatinine and metanephrines to creatinine best assesses children's 24-hour urine tests.

   Commonly, the adult plasma free metanephrine ULN is reported to be 57 pg/mL (0.31 nmol/L), with the adult plasma free normetanephrine ULN reported to be 148 pg/mL (0.75 nmol/L). However, reference ranges vary greatly between laboratories. The reported ULN for plasma free metanephrine ranges from 55 pg/mL to 85.6 pg/mL (0.30-0.47 nmol/L), and the reported ULN for plasma free normetanephrine ranges from 110 to 200 pg/mL (0.56-1.02 nmol/L). Plasma norepinephrine and normetanephrine levels increase with advancing age. For patients over age 65, the plasma-fractionated free normetanephrine ULN should be considered 200 pg/mL (1.02 nmol/L).
3. **Medical conditions** On average, hospitalized patients and those with essential hypertension have higher levels of catecholamines and metanephrines (plasma and urine) than do matched nonhospitalized and normotensive individuals. Therefore, many laboratories have separate reference ranges for hypertensives and nonhypertensives.

Patients with illness, trauma, or sleep apnea have increased excretion of both catecholamines and metanephrines. Patients with renal failure on dialysis have elevated levels of plasma catecholamines (58%), plasma free metanephrines (25%), and plasma total (deconjugated) metanephrines (100%). Patients with partial renal insufficiency also have misleadingly elevated levels of plasma catecholamines (32%), plasma free metanephrines (26%), and plasma total metanephrines (50%). Thus, in patients with renal failure, the best screening test is plasma free metanephrines, but the test still lacks specificity when elevated. Serum CgA levels are also elevated in renal insufficiency.

Misleading elevations of at least one metanephrine or catecholamine determination occur in 10% to 20% of tested individuals without PHEO/PGL. These elevations are typically less than 50% above the maximum normal and often normalize on retesting. Patients with PHEOs usually have elevations of metanephrines or catecholamines that are more than three times normal. In one large series, a false-positive elevation in at least one test occurred

in 22% and marked elevations in at least one test occurred in 3.5% of patients with no PHEO. False-positive test results were judged to have occurred from physiologic variation (33%), laboratory errors (29%), or drug interference (21%).

Plasma normetanephrines reflect disease activity in patients with secretory PGLs or metastases. Normal ranges for plasma metanephrines in children are different from those of adults and have been reported by Weise et al, 2002. Plasma normetanephrine levels typically increase with age; about 16% of older patients being evaluated for PHEO/PGL have levels above the published normal range for young adults (false positives); only 3% of young adults have false-positive plasma normetanephrine concentrations. Stimulation tests are not recommended (discussed later).

A. Metanephrines and catecholamines

Plasma-fractionated free metanephrines The most sensitive and simple screening test for PHEO/PGL is an assay for plasma-fractionated (metanephrine and normetanephrine) free metanephrines. This assay is particularly useful to screen patients for secretory PHEO/PGL (see Table 11–7). For younger children, plasma testing is particularly preferred over urine testing due to the relative ease of collection. This assay's sensitivity pertains to secretory PHEO/PGL. However, some tumors, particularly PGLs, may not secrete any metanephrines. So to screen for a PGL in a normotensive patient with a suspicious mass, it is recommended that plasma dopamine and serum CgA also be tested, since some PGLs secrete purely dopamine or only CgA (discussed later).

In the United States, most reference laboratory assays for metanephrines employ high-pressure liquid chromatography (HPLC) with tandem mass spectrometry (MS/MS). This test is 97% sensitive. The 3% of affected patients with PHEO/PGL with normal plasma-fractionated free metanephrines usually have very small tumors, nonsecreting tumors, or dopamine-secreting tumors such that they are not hypertensive. Although MS/MS reduces drug interference that can be a problem with assays that employ HPLC with electrochemical detection (ECD), the test is only 87% specific for PHEO/PGL in a referred patient population.

Patient preparation: In order to reduce the chance of false-positive testing, patients having plasma-fractionated free metanephrines determinations by tandem mass spectroscopy (MS/MS) should observe the following precautions: No decongestants, amphetamines, MAO inhibitors, or cocaine for at least 7 days before testing. The patient must not have any current stressful illness or be in withdrawal from alcohol, narcotics, or clonidine. There must be full recovery from any postoperative pain. The patient should have no local anesthetics (lidocaine) or acetaminophen for 24 hours before blood draw. Overnight fasting is preferred (water permitted). Patients should avoid coffee (including decaffeinated coffee) tea, tobacco, and strenuous exercise on the day of the blood draw. The patients being screened should be relaxed and ideally resting supine for at least 15 minutes prior to the blood draw; if that is not possible in an outpatient laboratory, the patient should at least sit quietly for a minimum of 15 minutes before the blood draw.

Since the test's specificity is much lower than its sensitivity, most patients with levels above the reference range do not harbor

a tumor. This is particularly true when plasma levels are less than three times higher than the upper limit of the reference range. Conversely, it is extremely likely that a patient harbors a PHEO when the plasma free metanephrine is above 220 pg/mL (1.2 nmol/L) or the plasma free normetanephrine is above 430 pg/mL (2.2 nmol/L). Patients with abnormally high plasma-fractionated metanephrines usually need to have the test repeated under controlled conditions (see Patient Preparation, discussed earlier), along with a 24-hour urine collection for fractionated metanephrines, catecholamines, and creatinine (discussed later).

Normally, about 90% of circulating metanephrine and about 50% of circulating normetanephrine originate directly from the adrenal medulla. The term **total metanephrines** refers to both normetanephrine and metanephrine. There are two circulating forms of normetanephrine and metanephrines: free and sulfate-conjugated metanephrines. The free metanephrines produced in the adrenal medulla and paraganglia are sulfate-conjugated by intestinal tissue; the sulfated form represents 97% of circulating metanephrines. Only 3% of total circulating metanephrine is free. Plasma metanephrine levels are sometimes measured after a deconjugation step, such that both free and conjugated metanephrines are measured; this assay is termed **deconjugated metanephrines** and reflects mostly the sulfate-conjugated species. The plasma free metanephrines assay is superior to the total (deconjugated) metanephrine assay.

The sensitivity of the plasma free metanephrines is due to the tumor's continuous secretion of metanephrines. Catecholamines in secretory vesicles exist in a dynamic equilibrium with the surrounding cytoplasm, with catecholamine uptake into the vesicles being balanced by their leakage into the cytoplasm. In the cytoplasm, the enzyme COMT converts epinephrine to metanephrine and converts norepinephrine to normetanephrine. The catecholamine metabolites then leak out of the cell continuously to become free metanephrines. While catecholamines are secreted in bursts associated with exocytosis of neurosecretory vesicles, free metanephrines are produced continuously. This eliminates the need to catch a paroxysmal hypertensive event. Plasma free metanephrine levels are within the reference range in 75% of patients on dialysis and 74% of patients with renal insufficiency. In contrast, plasma deconjugated metanephrines are within the reference range in 0% of patients on dialysis and in 50% of patients with renal insufficiency.

Factors causing misleading plasma free fractionated metanephrine levels (Table 11–15) *Smoking* can elevate plasma free metanephrine assays. The patient must not smoke for at least 4 hours before collection. *Coffee* increases plasma normetanephrine by 20% and *food consumption* increases it by 8%; but they have no effect upon plasma metanephrine levels. *Sympathomimetics* such as decongestants and amphetamine derivatives should be discontinued for at least 7 days before the test. Other drugs that can misleading elevate plasma free metanephrines include *local anesthetics, cocaine, lidocaine, halothane anesthesia, and MAO inhibitors. Acetaminophen is one of the few medications that can actually interfere with the MS/MS assay.* Conditions that increase plasma metanephrine and normetanephrine levels include

any *severe physical stress such as acute illness or narcotic or alcohol withdrawal, anxiety, sleep apnea, renal failure, and essential hypertension. Physical exercise* raises plasma free metanephrine and normetanephrine by over 80%. There can be seasonal variations in plasma normetanephrine levels, with 20% higher levels during *winter* being reported in Germany and the Netherlands.

Assumption of a supine resting position reduces plasma metanephrine by an average of 34% and reduces normetanephrine by 19%, compared to standing at rest. Blood samples should ideally be drawn from an indwelling heparin-locked intravenous line while the patient is fasting and resting supine. However, it is not usually practical for patients to rest supine for prolonged periods in outpatient laboratories, so patients are advised to sit quietly for at least 15 minutes prior to the blood draw. Blood should be centrifuged immediately and stored at 4°C to improve stability.

In patients with adrenal PHEO, over 90% of circulating metanephrine originates from the tumor, while variable percentages of circulating normetanephrine originate from the tumor. Patients with elevated plasma metanephrine or epinephrine are likely to have an adrenal PHEO, since PGLs do not ordinarily secrete epinephrine or metanephrine. Metastases from epinephrine-secreting adrenal PHEOs may sometimes continue to secrete epinephrine and its metabolite metanephrine.

24-Hour urine-fractionated metanephrines

This assay, which measures the sum of both conjugated and free urinary metanephrines, rivals plasma free metanephrines in sensitivity. In the United States, most reference laboratories employ tandem mass spectrometry (MS/MS), which eliminates interference from most drugs and foods. Urinary metanephrines are rather stable compounds, so it is not necessary to collect the specimen with acid preservative. Conversely, acid preservative (used for collection of urinary catecholamines) does not adversely affect metanephrines. Urine fractionated 24-hour metanephrines are useful to confirm high plasma levels or to help rule-out a PHEO/PGL in patients with marginal elevations in plasma-fractionated metanephrines. The disadvantages of 24-hour urine collections include: (1) the inconvenience to the patient; (2) the likelihood that there will be an error in the patient's collection or in the laboratory's handling of the specimen; (3) the lower specificity for urinary fractionated metanephrines (69%) compared to plasma-fractionated free metanephrines (89%); (4) the inaccuracy for patients with renal failure (see Appendix for normal reference ranges for 24-hour metanephrines).

Like plasma-fractionated metanephrines, urinary metanephrines are much more sensitive than they are specific. Most patients with elevated levels of fractionated metanephrines do not harbor a PHEO if their levels are less than three times the upper limit of the reference range. Conversely, patients are very likely to have a pheochromocytoma if the 24-hour urine metanephrine excretion is above 600 μg/d (3.0 μmol/d) or normetanephrine excretion is above 1500 μg/d (8.2 μmol/d).

As with plasma metanephrines, patients with a definitely elevated urinary metanephrine are likely to harbor an adrenal PHEO, since PGLs rarely secrete epinephrine or metanephrine.

Plasma-fractionated catecholamines

Misleading elevations in plasma catecholamines commonly occur from the stress of having the phlebotomy. Normal plasma levels of catecholamines are listed in the Appendix. In patients with PHEO/PGL, plasma concentrations of norepinephrine do not correlate well with blood pressure.

Although plasma metanephrines are more sensitive than plasma catecholamines for detecting a PHEO, plasma catecholamines are of value in helping distinguish true-positive from false-positive results. In normal individuals, about 100% of circulating epinephrine originates from the adrenal medulla, while over 90% of circulating norepinephrine originates from peripheral sympathetic synapses. In both normals and those with PHEOs, plasma norepinephrine levels fluctuate with the degree of peripheral sympathetic activation and both epinephrine and norepinephrine levels can increase during PHEO paroxysms. In patients with PHEOs, the tumor's production of metanephrine metabolites is rather constant and relatively unrelated to peripheral sympathetic activity or a tumor's paroxysmal catecholamine secretion. Patients with false-positive test results due to sympathetic activation tend to have the following pattern: a higher percentage increase of plasma norepinephrine above the upper limit of the reference range compared to plasma normetanephrine; a higher percentage increase of plasma epinephrine above the upper limit of the reference range compared to plasma metanephrine. Plasma dopamine should be assayed in normotensive patients with a mass that is suspicious for a PGL, since some PGLs secrete only dopamine.

Plasma catecholamines are of limited value in patients on dialysis and in those with renal insufficiency. About 58% of dialysis patients have plasma catecholamines above the normal reference range; about 32% of renal insufficiency patients have catecholamines above the normal reference range.

Most assays for plasma catecholamines currently employ HPLC with ECD; assays using tandem mass spectrometry (MS/MS) have not become generally available.

Urine-fractionated catecholamines and dopamine

Adult normal maximal urinary concentrations for catecholamines and their metabolites are listed in the *Appendix*. A single 24-hour urine specimen is collected for fractionated catecholamines, fractionated metanephrines, dopamine, and creatinine. The container is acidified with 10 to 25 mL of 6 N HCl for preservation of the catecholamines; the acid does not interfere with metanephrine and creatinine assays. The acid preservative may be omitted for children for safety reasons, in which case the specimen should be kept cold and processed immediately. The laboratory requisition form should request that assays for fractionated catecholamines, fractionated metanephrines, and creatinine be performed on the same specimen. Urinary dopamine determination is not a sensitive test for PHEOs. However, in patients with established PHEOs, a normal urine dopamine is fairly predictive of benignity, whereas elevated urine dopamine excretion is seen in both benign and malignant PHEOs.

Single-void (spot) urine specimens

can be collected on first morning void or following a paroxysm. No acid preservative is used on single-void specimens, because it dilutes the specimen and

is not required. For single-void collections, patients are instructed to void and discard the urine immediately at the onset of a paroxysm and then collect the next voided urine. The laboratory requisition should request spot urine for total metanephrines and creatinine concentrations. It is prudent to contact the laboratory directly and explain that the specimen is meant to be a single-void urine and not a 24-hour specimen, else the specimen may be rejected. Patients with PHEOs generally excrete over 2.2 μg total metanephrine/mg creatinine.

B. Serum CgA   Some tumors (mostly PGLs) fail to secrete catecholamines or metanephrines. Serum CgA is a useful test to diagnose and monitor such nonsecretory tumors. CgA may be determined by immunoradiometric assays. The serum CgA assay has become useful for the diagnosis of PHEO. However, CgA undergoes extensive tumor-specific cleavages so that only certain serum assays are useful for clinical diagnosis.

Serum CgA levels have a circadian rhythm in normal individuals, with lowest levels found at 8 AM and higher levels in the afternoon and at 11 PM. CgA levels are not elevated in essential hypertension. CgA is also secreted from extra-adrenal sympathetic nerves. Serum CgA levels are elevated in the great majority of patients with PHEOs. The serum levels of CgA correlate with tumor mass, making CgA a useful tumor marker. However, smaller tumors may not be diagnosed. Serum CgA levels tend to be particularly elevated in patients with metastatic PHEO.

Serum CgA can be elevated even in patients with biochemically silent tumors. In patients with normal renal function, serum CgA has a sensitivity of 83% to 90% and a specificity of 96% for diagnosis of these tumors. However, the usefulness of serum CgA levels is negated by any degree of renal failure because of its excretion by the kidneys; even mild azotemia causes serum levels to be elevated. However, in patients with normal renal function, a high serum level of CgA along with high urine or plasma catecholamines or metanephrines is virtually diagnostic of PHEO.

Since CgA is also cosecreted with gastrin, serum CgA levels are also elevated in conditions with elevated serum gastrin: atrophic gastritis, pernicious anemia, postvagotomy, gastrinoma, gastric carcinoma, carcinoid tumor, and small cell lung carcinoma. CgA levels are elevated in about 60% of patients taking proton pump inhibitors (PPIs), but not H_2 blockers. Serum CgA levels rise variably after meals, so blood for CgA should be drawn after an overnight fast or repeated fasting if a nonfasting level is elevated. False-positive testing has also occurred in patients with inflammatory bowel disease, liver disease, hepatocellular carcinoma, prostate cancer, pituitary tumors, rheumatoid arthritis, and stress. False-positives have been reported to be due to heterophile antibody interference with the assay. Serum CgA levels can also be elevated without any discernable cause.

C. Suppression and stimulation testing   Glucagon stimulation testing is no longer used since it can cause dangerous hypertension.

Clonidine suppression test: This test may help distinguish patients with PHEO from normals with elevated normetanephrine levels. Clonidine is a central α_2 adrenergic blocker that suppresses the release of norepinephrine at sympathetic nerve synapses, thereby reducing circulating levels of norepinephrine and its metabolite, normetanephrine. In contrast, in patients with PHEO, most circulating normetanephrine is derived from continuous leakage from the tumor, such that clonidine is less able to suppress it. This test is most accurate when free normetanephrine is assayed, rather than norepinephrine.

To perform the clonidine suppression test, the patient must be fasting overnight and avoid smoking or interfering medications such as phenoxybenzamine, β-adrenergic blockers, tricyclic antidepressants and diuretics for at least 48 hours. An indwelling venous catheter is inserted, and the patient should remain recumbent. Thirty minutes later, blood is drawn for baseline plasma free normetanephrine. Clonidine is then given orally in a dose of 0.3 mg/kg; 3 hours afterward, blood is again drawn for plasma free normetanephrine.

In a study of 49 normals without PHEO, clonidine suppressed plasma normetanephrine levels more than 40% or to below 112 pg/mL in 100%. However, clonidine failed to suppress normetanephrine in 46 of 48 patients with PHEO. Despite the potential helpfulness of the clonidine suppression test, it cannot be completely relied upon. Unnecessary surgeries have been performed on the basis of misleading clonidine suppression testing.

D. Other laboratory tests

1. **Urine VMA** Urinary VMA determinations have an overall diagnostic sensitivity for PHEO of only about 63% and do not improve the sensitivity or specificity of other tests for the diagnosis of PHEO. However, some centers have traditionally used a combination of 24-hour urinary VMA, metanephrine, and creatinine determinations with good results. VMA is stable for 5 days at room temperature; 6 N HCl is used to preserve urine specimens that are stored for longer than 5 days before analysis. Before urine collections for VMA testing, patients must avoid salicylates, caffeine, phenothiazines, and antihypertensive drugs for 72 hours. Coffee, tea, chocolate, bananas, and vanilla must also be avoided. Normal ranges for VMA vary by age (see Appendix).

2. **Plasma renin activity** Levels of plasma renin activity are not typically suppressed in patients with PHEO/PGL because catecholamines stimulate renin release, and some tumors may secrete renin ectopically.

3. **Other tests** Patients with PHEO/PGL are frequently found to have an increased white blood count with a high absolute neutrophil count. Counts as high as 23,600/μL have been reported. Marked eosinophilia may sometimes occur. Hyperglycemia is noted in about 35% of patients with PHEO, but frank diabetes mellitus is uncommon. The erythrocyte sedimentation rate is elevated in some patients. Hypercalcemia is common and may be caused by bone metastases or tumoral secretion of PTHrP. Hypocalcemia occurs rarely. Erythrocytosis sometimes occurs, usually due to volume contraction and rarely due to ectopic secretion of erythropoietin.

Factors That May Cause Misleading Biochemical Testing for PHEO (Table 11–15)

Several different methods may be employed for assay of urine and plasma catecholamines and metanephrines. Each assay uses different

TABLE 11–15 Factors that can cause misleading catecholamine or metanephrine results.[a]

Drugs	Foods	Conditions
Acetaminophen[b]	Bananas[c]	Age (children & elderly)[f]
Aldomet[b]	Caffeine[c]	
Amphetamines[c]	Coffee[b]	Amyotrophic lateral sclerosis[c]
Anesthetics (local)[f]	Curry leaves[b]	
Bronchodilators[c]	Peppers[b]	Brain lesions[c]
Buspirone[b]	Pineapples[c]	Anxiety[c,f]
Captopril[b]	Walnuts[c]	Carcinoid[c]
Cimetidine[b]		Drug withdrawal (narcotic, alcohol, clonidine)[c,f]
Cocaine[c]		
Codeine[b]		
Contrast media (meglumine acetrizoate, meglumine diatrizoate)[e]		Eclampsia[c]
		Emotion (severe)[c,f]
		Essential hypertension[c]
Decongestants[c]		Exercise (vigorous)[c,f]
Ephedrine & Epinephrine[c]		Guillain-Barré syndrome[c]
Halothane anesthesia[f]		Hypoglycemia[c]
Fenfluramine[d]		Lead poisoning[c]
Isoproterenol[c]		Myocardial infarction (acute)[c]
Labetalol[b,c]		
Levodopa[b]		Pain (severe)[c]
Mandelamine[b]		Porphyria (acute)[c]
Mesalamine[b]		Psychosis (acute)[c]
Methenamine[d]		Quadriplegia
Metoclopramide[b]		Renal failure[c,d,f]
Monoamine oxidase inhibitors[f]		Sleep apnea[c,f]
		Winter[f]
Nitroglycerin[c]		
Phenoxybenzamine[b,c]		
Serotonin-norepinephrine reuptake inhibitors (SSNRI): venlafaxine, desvenlafaxine, duloxetine, levomilnacipran[c]		
Smoking[f]		
Sulfasalazine[b]		
Tricyclic antidepressants[b,c]		

[a]Note that assays for metanephrine employing tandem mass spectroscopy (MS/MS) are not prone to interference from drugs or foods, except those that increase catecholamine excretion. Reproduced with permission from Papadakis MA, McPhee SJ, Rabow MW. *Current Medical Diagnosis & Treatment 2017*. New York: McGraw Hill Education; 2017.

[b]May cause confounding peaks for catecholamines with an HPLC-ECD chromatogram.

[c]Increases catecholamine excretion.

[d]Decreases catecholamine excretion.

[e]May decrease urine metanephrine excretion.

[f]Increases plasma and urinary normetanephrine and metanephrine.

methods and internal standards. Most assays for catecholamines now employ HPLC with ECD. Such assays can be affected by interference from a diverse range of drugs and foods. These substances cause unusual shapes in the peaks on the chromatogram. Not all of these assays are the same, and the potential for interference depends on the particular method employed. Therefore, it is best to check with the reference laboratory that runs the test or provides the test kit.

A. Drugs   Certain radiopaque contrast media, including those that contain meglumine acetrizoate or meglumine diatrizoate (eg, Renografin, Hypaque-M, Renovist, Cardiografin, Urografin, and Conray), can falsely lower urinary metanephrine determinations in some assays for up to 12 hours following administration. However, diatrizoate sodium is an intravenous contrast agent that does not cause such interference and should be requested if a CT scan must be performed prior to testing for metanephrines. Many other drugs cause interference in the older fluorometric assays for VMA and metanephrines. Tandem mass spectrometry for metanephrines virtually eliminates direct drug interference in the assay.

B. Foods   Even while using HPLC-ECD assay techniques, certain foods can cause misleading results in assays for catecholamines and metanephrines (see Table 11–15). Coffee (even if decaffeinated) contains substances that can be converted into a catechol metabolite (dihydrocaffeic acid) that may cause confusing peaks on HPLC. Caffeine inhibits the action of adenosine; one action of adenosine is to inhibit the release of catecholamines. Heavy caffeine consumption causes a persistent elevation in norepinephrine production and raises blood pressure an average of 4 mm Hg systolic. Bananas contain considerable amounts of tyrosine, which can be converted to dopamine by the central nervous system; dopamine is then converted to epinephrine and norepinephrine. Dietary peppers contain 3-methoxy-4-hydroxybenzylamine (MHBA), a compound that can interfere with the internal standard used in some assays for metanephrines.

C. Diseases   Any severe stress can elicit increased production of catecholamines and metanephrines. Urinary excretion of catecholamines and metanephrines is reduced in renal failure.

Differential Diagnosis of Pheochromocytoma and Paraganglioma (Table 11–16)

PHEOs and PGLs have such protean manifestations that many conditions enter into the differential diagnosis. Essential hypertension is extremely common, and it is not practical to screen for PHEO in all patients with elevated blood pressure. However, PHEO should enter the differential diagnosis for any hypertensive patient having blood pressures above 180 mm Hg systolic and for any hypertensive patient who has one of the following symptoms: headaches, palpitations, sweating episodes, or unexplained bouts of abdominal or chest pains.

Anxiety (panic) attacks begin abruptly and can be associated with tachycardia, tachypnea, and chest discomfort, symptoms that are commonly seen with PHEOs. However, patients with panic attacks are more likely to have a precipitating social situation, tend to be exhausted for more than 2 hours following an attack, live in dread of the next attack, and often change their activities to avoid situations that might trigger anxiety.

Renal artery stenosis and renal parenchymal disease can cause increased secretion of renin resulting in severe hypertension. However, a detectable serum renin level does not exclude PHEO,

TABLE 11–16 Differential diagnosis of pheochromocytoma.

Acute intermittent porphyria
Autonomic epilepsy
Cardiac arrhythmias
Clonidine withdrawal
Coronary vasospasm
Encephalitis
Erythromelalgia
Essential hypertension
Hypertensive crisis associated with:
Cerebrovascular accidents
Surgery
Acute pulmonary edema
Severe pain
Hypertensive crisis of MAO inhibitors
Hypoglycemia
Hypogonadal hot flushes
Lead poisoning
Mastocytosis
Migraine and cluster headache
Renovascular hypertension
Sleep apnea
Sympathomimetic drug ingestion
Thyrotoxicosis
Toxemia (eclampsia) of pregnancy

because catecholamines can stimulate renin secretion, and PHEOs can secrete renin ectopically. Furthermore, large PHEOs and PGLs arising near the renal hilum can occlude the renal artery, causing concomitant renovascular hypertension.

Hypogonadism can cause vasomotor instability in both women and men; attacks of flushing, sweating, and palpitations can mimic symptoms seen with PHEO. Factitious symptoms may be caused by surreptitious self-administration of various drugs such as epinephrine. Hyperthyroidism can cause heat intolerance, sweating, palpitations, and systolic hypertension with a widened pulse pressure. Carcinoid syndrome causes flushing during attacks but usually without pallor, hypertension, palpitations, or diaphoresis.

Obstructive sleep apnea can cause systemic hypertension; recurrent nocturnal hypoxia results in repeated episodes of stressful arousal that cause bursts of secretion of catecholamines, particularly epinephrine. Sleep apnea has been reported to cause misleading increases in the urinary excretion of catecholamines.

Patients with erythromelalgia can have episodic hypertension, but it is associated with flushing of the face and legs during the attack; patients with PHEO typically have facial pallor during attacks. With erythromelalgia, patients have painful erythema and swelling in the legs that is relieved by application of ice; such symptoms are not characteristic of PHEO.

Patients who have intermittent bizarre symptoms may have their blood pressure and pulse checked during a symptomatic episode with a home blood pressure meter or an ambulatory blood pressure monitor. Those who are normotensive during an attack are not likely to harbor a PHEO.

PHEOs and PGLs often present with abdominal pain and vomiting. Such symptoms are similar to those of an intra-abdominal emergency, particularly in the presence of leukocytosis and fever,

which can also be seen with PHEOs. Abdominal pain usually prompts a CT scan of the abdomen, which generally shows the PHEO or PGL. Even after detection on CT scan, PHEOs and juxtarenal PGLs may be mistaken for renal carcinoma. Large left-sided PHEOs are often mistaken for carcinoma of the tail of the pancreas.

Neuroblastoma is the most common extra-cranial solid malignancy of childhood. It is also the most common malignancy in infants under 18 months old. Neuroblastomas are derived from the embryonic neural crest of the peripheral sympathetic nervous system. They most often arise in the adrenal gland, but can also develop in sympathetic nerve ganglia near the cervical or thoracic vertebrae or in the pelvis. They metastasize to bones, lymph nodes, liver, and skin. Symptoms depend upon the age of the patient, site of origin, degree of metastatic involvement, and the systemic response to the tumor. Infants usually present with localized disease (Stage 1 or 2) or with a special disseminated pattern of disease (Stage 4S—infants with metastatic disease to liver and/or skin) associated with a favorable outcome. In contrast, most children over 1 year of age present with advanced disease (Stage 3 or 4). Unfavorable biologic features include amplification of myelomatosis viral-related oncogene, neuroblastoma (MYCN), deletion or loss of heterozygosity of chromosome 1p or 11q and gains at 17q. Favorable biologic features include hyperdiploidy and overexpression of the gene encoding the nerve growth factor receptor, tyrosine kinase A (TrkA). About 85% of affected children secrete excessive catecholamines—but rarely in sufficient amounts to cause symptomatic hypertension or the paroxysms typical of PHEOs. Most neuroblastomas concentrate ^{123}I-MIBG but they can be distinguished from PHEOs and PGLs by clinical and histologic criteria.

Localization Studies for Pheochromocytoma

There are several available imaging modalities for PHEO/PGL, each having unique sensitivity and specificity. There is no single imaging study that is 100% sensitive and specific. Generally, when a PHEO or PGL is suspected, the initial diagnosis is best made biochemically. The initial localization scan is with CT or MRI of the abdomen and pelvis, which will detect about 95% of these tumors. However, since PGLs can arise in the chest, further scanning of the chest may be required. Diagnostic confirmation that the tumor is a PHEO or PGL may be done with either ^{123}I-MIBG or ^{18}F-FDA positron emission tomography (PET) scanning; unfortunately, these scans are only 78% sensitive for these tumors. Scanning with ^{18}F-FDG or ^{18}F-FDA PET is more sensitive for detecting metastases than is ^{123}I-MIBG.

A. Computed tomography (Figure 11–11)   After a PHEO has been diagnosed by clinical and biochemical criteria, hypertension must first be controlled (discussed later), because intravenous contrast can precipitate a hypertensive crisis. The PHEO must then be localized. It is useful to perform an initial nonenhanced (without intravenous contrast) CT of the adrenals because the density of an adrenal tumor can be better approximated without
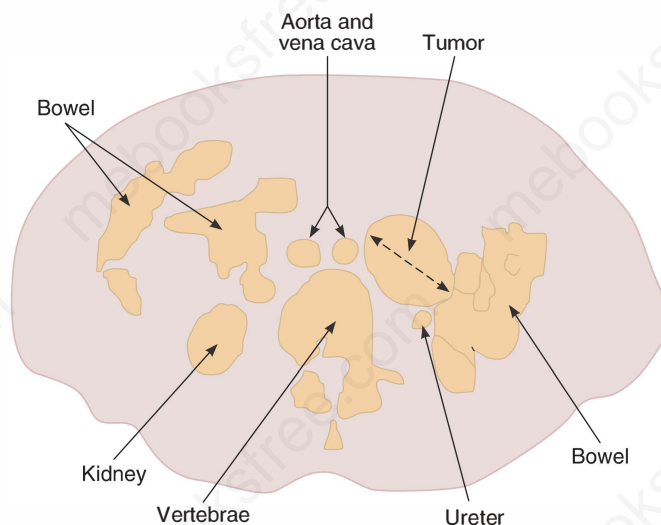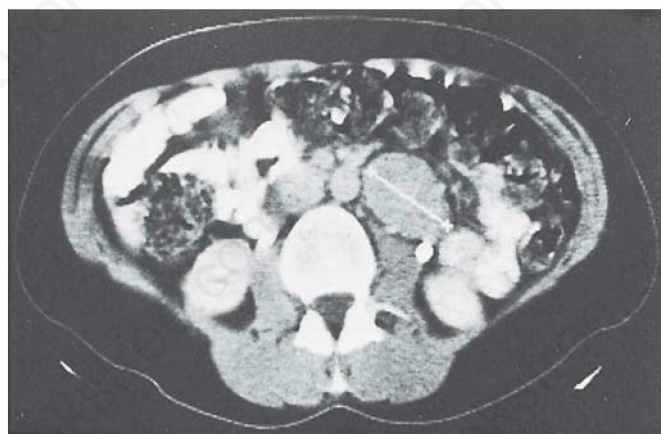
FIGURE 11–11 Left infrarenal PGL shown by CT scanning. The diagram on the right identifies many of the visible structures.

intravenous contrast. An adrenal mass with a density of less than 10 Hounsfield units (HU) is unlikely to be a PHEO. CT scanning of the entire abdomen (from the diaphragm through the pelvis) is obtained with intravenous nonionic contrast-enhanced and delayed contrast-enhanced imaging. Thin-section (2-5 mm) cuts should be obtained through the adrenals with an adrenal protocol that specifically looks for a vascular tumor blush and determines rate of contrast washout. PHEOs tend to have a tumor blush and slower contrast washout than adrenocortical adenomas. Hypertensive crises have been provoked in patients with PHEO/PGL who receive intravenous contrast. Ionic intravenous contrast frequently triggers hypertension, while nonionic intravenous contrast infrequently triggers hypertension. Therefore, all patients with suspected PHEO/PGL should have hypertension controlled before receiving intravenous CT contrast, and nonionic contrast is strongly preferred. Care must be taken in patients with contrast allergy, since glucocorticoids that are typically given by protocol to such patients can precipitate hypertensive crisis in patients with secretory PHEO/PGL. Glucagon should not be used during a CT because it may provoke a hypertensive crisis.

If no mass is discovered, ^{123}I-MIBG imaging may be obtained or the CT scan may be extended into the chest in search of a PGL—or both procedures may be employed. The great majority of PHEOs are over 2 cm in diameter, well within the resolution capacity of the CT scan.

The overall sensitivity of CT scanning for an adrenal PHEO is about 90%—and over 95% for PHEOs that are over 0.5 cm in diameter. However, CT scanning is less sensitive for the detection of small adrenal PHEOs or adrenal medullary hyperplasia; this becomes an important issue in patients with MEN 2 or VHL disease. CT is also less sensitive for detecting extra-adrenal PGLs and early recurrent tumors in the adrenal surgical bed. CT will not detect small metastases or some metastases that strictly involve the bone marrow without osteoclastic activity. Metal surgical clips pose problem for CT scanning, causing distortion artifact and reducing the resolution of the scan.

B. Magnetic resonance imaging   MRI is useful in the diagnosis of adrenal PHEOs, PGLs, and metastatic disease. It may be used with or without gadolinium contrast. However, intravenous gadolinium contrast does not cause hypertensive crisis and improves resolution for metastases, particularly hepatic metastases. MRI scanning with intravenous gadolinium contrast is also useful for patients with a known allergy to intravenous iodinated CT contrast agents. MRI is the scanning technique of choice in children and during pregnancy, because it involves no radiation exposure. Since MRI scanning delivers no radiation, it is preferred for serially scanning patients known to harbor a gene mutation predisposing them to PHEO/PGL.

MRI can help determine whether an adrenal mass is a PHEO when biochemical studies are inconclusive. On MRI T1-weighted images, PHEOs have a dull signal (due to lack of fat cells), similar to kidney and muscle, distinguishing it from adrenal cortical adenomas, which contain fat and therefore have an intensely bright signal on T1-weighted images. The hypervascularity of PHEOs makes them appear bright on MRI T2-weighted images, without signal loss on opposed phase imaging. However, other adrenal malignancies, adrenal adenomas, and hemorrhages can also appear bright on MRI T2-weighted images. Therefore, MRI scanning cannot definitively identify an adrenal mass as a PHEO. MRI of the abdomen has a sensitivity of about 95% for adrenal PHEOs over 0.5 cm in diameter. Like CT scanning, MRI is less sensitive for the detection of extra-adrenal PGLs, metastatic disease, and recurrent small tumors in the adrenal surgical bed. MRI is helpful in visualizing PGLs that are intracardiac, juxtacardiac, or juxtavascular; MRI is particularly important for patients with PGLs adjacent to the vena cava or renal vein to detect vascular invasion. MRI is superior to CT in visualizing PGLs of the bladder wall. MRI can visualize some metastases to bone suspected on ^{123}I-MIBG imaging or PET scanning. Another advantage of MRI scanning is that retained internal metallic surgical clips do not cause the distorting reflection artifacts that occur with CT scanning.

The disadvantages of MRI scanning include expense and its inability to crisply image lungs (due to movement artifact).

Also, morbidly obese patients may not be able to fit into a standard helical MRI scanner; open MRI scanners can be used but are generally less sensitive. Claustrophobic patients require a sedative before the scan or an open MRI. Patients with internal pacemakers or defibrillators may not have MRI scans; nor may patients with implanted neural stimulators, cochlear implants, Swan-Ganz catheters, insulin pumps, cerebral aneurysm clips, ocular metallic foreign bodies, or retained metal shrapnel or bullets. Patients with retained surgical clips, artificial heart valves and joints may have MRI imaging; the spine may be imaged in patients with spinal hardware, but the imaging can be distorted.

C. Metaiodobenzylguanidine scanning (Figure 11–12 and Table 11–17)

MIBG is a guanidine derivative that resembles norepinephrine and is actively transported into adrenal medullary cells via the norepinephrine transporter system, selectively accumulating in neurosecretory granules. Unlike norepinephrine, MIBG has low affinity for catecholamine receptors and is not metabolized.

Scintigraphy using ^{123}I-MIBG or ^{131}I-MIBG is useful for determining whether an adrenal mass is a PHEO, for imaging occult PGLs, and for confirming whether a certain extra-adrenal mass is a PGL or neuroblastoma. MIBG scanning is also useful for screening patients for metastases. MIBG uptake can often be seen in apparently nonfunctioning PHEOs.

The isotope that is preferable for precise imaging is ^{123}I-MIBG, because ^{123}I emits γ radiation at lower energy (159 keV) than does ^{131}I (364 keV). The lower-energy γ emissions allow crisper images

FIGURE 11–12 ^{123}I-metaiodobenzylguanidine (^{123}I-MIBG) scan of a woman with a large left PHEO. Normal ^{123}I-MIBG uptake is seen in the liver, salivary glands, and heart. ^{123}I-MIBG is renally excreted and is visible in the bladder.

TABLE 11–17 Drugs that potentially inhibit MIBG uptake by pheochromocytomas and paragangliomas. Inhibition can persist up to 2 weeks.

Antidepressant inhibitors of norepinephrine reuptake: bupropion, clomipramine, desvenlafaxine, duloxetine, levomilnacipran, mirtazapine, tricyclic drugs, venlafaxine
Antiemetics: metoclopramide, prochlorperazine, promethazine
Antipsychotics: phenothiazines, haloperidol, thiothixene
Amphetamines: amphetamine, dextroamphetamine, lisdexamfetamine, methamphetamine
Anorexic diet medications: benzphetamine, diethylpropion, phendimetrazine, phentermine, sibutramine
Antihypertensive: labetalol (2-6 weeks)
CNS stimulants: cocaine, methylphenidate, dexmethylphenidate
Decongestants (oral, nasal, ophthalmic): pseudoephedrine, ephedrine, phenylpropanolamine, naphazoline, oxymetazoline, propylhexedrine, tetrahydrozoline, xylometazoline
Herbal therapies: ephedra (ma huang), St. John's Wort, yohimbine
Monoamine oxidase inhibitors (MAOi): isocarboxazid, linezolid, phenelzine, selegiline, tranylcypromine

with ^{123}I-MIBG scanning as well as single photon emission computed tomography (SPECT). ^{123}I-MIBG SPECT scanning is more sensitive than ^{123}I-MIBG planar imaging for detecting small metastases and has the advantage of being able to do the scanning on the day following injection of the isotope. SPECT scanning with ^{123}I-MIBG can be combined with CT to produce a three-dimensional fusion scan; the resultant combined images can help distinguish whether a given mass has ^{123}I-MIBG uptake.

^{123}I-MIBG scanning has an overall sensitivity of 82% and a specificity of 82% for PHEOs and PGLs combined. The sensitivity of ^{123}I-MIBG scanning is 88% for primary adrenal PHEOs and 67% for PGLs, with an overall sensitivity of 78% for primary PHEO/PGL tumors. The sensitivity of ^{123}I-MIBG scanning for metastases is lower at only 57%. Scanning with ^{123}I-MIBG is more sensitive for PHEOs that are benign, unilateral, adrenal, capsule-invasive, and sporadic. Scanning with ^{123}I-MIBG is less sensitive for bilateral, malignant, extra-adrenal, noninvasive, and MEN 2-related or VHL-related PHEOs.

To block the thyroid's uptake of free radioiodine, saturated solution of potassium iodide, five drops orally three times daily, is given before the injection and daily for 4 days after ^{123}I-MIBG and for 7 days after ^{131}I-MIBG. Scanning is performed 24 to 48 hours after ^{123}I-MIBG infusion and 48 to 72 hours after ^{131}I-MIBG infusion.

False-negative MIBG scans are seen in about 15% of cases of either benign or malignant PHEO. False-negative scans can occur in patients who have taken certain drugs (eg, tricyclic antidepressants or cyclobenzaprine) within 6 weeks. Other drugs that can cause false-negative scans when taken within 2 weeks include amphetamines, phenylpropanolamine, haloperidol, phenothiazines, nasal decongestants, cocaine, and diet pills. Labetalol reduces MIBG uptake, but the scan can still be done, albeit with suboptimal sensitivity (Table 11–17). When plasma norepinephrine levels are over 500 pg/mL (3 nmol/L), cardiac visualization is reduced on ^{18}F-dopamine PET scanning. This phenomenon is believed due to competitive inhibition of uptake-1 by high levels

of circulating norepinephrine. Therefore, it is likely that very high endogenous norepinephrine levels may reduce the sensitivity of ^{123}I-MIBG scanning as well as PET scanning that employs ^{18}F-DA or ^{19}F-DOPA because these compounds are transported into PHEOs, PGLs, and their metastases by the same uptake mechanism.

False-positive MIBG scans occur infrequently. Following ^{123}I-MIBG, some uptake in a normal adrenal medulla is seen in 32% to 75% of patients at 24 hours. Following ^{131}I-MIBG, uptake in a normal adrenal medulla is seen in 16% of normals at 48 hours. Adrenal uptake is often asymmetric and can be misinterpreted as showing a tumor. MIBG has renal excretion, so the renal pelvis and bladder are usually visualized on scanning and must be distinguished from tumor. If a patient is being evaluated for a bladder mass (to exclude a PGL), a bladder catheter may be inserted and the bladder flushed with saline to distinguish tumor from renal excretion of isotope. Urine contamination with ^{123}I-MIBG can also cause a false-positive scan. False-positive results have been reported with adrenal carcinomas and infections such as actinomycosis. The salivary glands are typically visualized since they are richly innervated. The heart, liver, and spleen normally take up some ^{123}I-MIBG. Some isotope is excreted in the stool, and intracolonic collections can be mistaken for tumor. When there is doubt about whether an area of uptake is a tumor, scanning can be repeated the next day, preceded by a laxative if required.

D. Positron emission tomography (Figure 11–13)   PET scanning can be performed shortly after the intravenous infusion of short-lived isotopes that are tagged to a compound that is preferentially absorbed by tumor tissue.

PET scanning has certain advantages over MIBG scanning. PET scanning can be done almost immediately. This gives it some advantage over MIBG scanning, which must be delayed for 24 to 48 hours after the injection to allow dissipation of background radiation. PET scanning does not require pretreatment with iodine to protect the thyroid, as is necessary with MIBG scanning. However, PET scanning is very expensive. The isotope ^{18}F has a half-life of just 2 hours and must be produced in a cyclotron, so ^{18}F-FDG or ^{18}F-FDA PET scanning is practical only at medical centers that have a cyclotron nearby.

1. **PET scanning with ^{18}F-FDG** Deoxyglucose (DG) is absorbed by tissues with active metabolism, including tumors. For PET scanning, DG is tagged with ^{18}F to produce fluorodeoxyglucose, or ^{18}F-FDG. This scan is widely available as a fusion scan with CT, ^{18}F-FDG PET/CT. This scan's sensitivity is 88% for nonmetastatic PHEO/PGL and 76% for metastatic PHEO/PGL. The sensitivity of ^{18}F-FDG-PET/CT is lower for indolent tumors and for patients who are diabetic or who are not fasting before the scan. ^{18}F-FDG-PET localizes in other tissues with a high metabolic rate, including areas of inflammation, brown fat, shivering muscles, but such areas of PET uptake can be distinguished from tumors by the lack of a CT correlate. However, ^{18}F-FDG-PET/CT detects other tumors besides PHEO/PGL and is, therefore, less specific for PHEO/PGL than is ^{18}F-FDA-PET or ^{123}I-MIBG scanning. ^{18}F-FDG/CT scanning is usually performed without CT contrast, such that

FIGURE 11–13 ^{18}F-deoxyglucose (^{18}FDG) PET/CT fusion scan. Patient had a retroperitoneal PGL that was resected. ^{18}FDG-PET/CT fusion scan shows a large metastasis involving the left acetabulum and inferior pubic ramus. The metastasis had negligible ^{123}I-MIBG uptake. A metastasis causing L4 compression had been treated previously with radiation therapy.

it misses certain tumors, particularly indolent PGL/PHEO metastases to liver, where the hepatic metastases may be invisible on both ^{18}F-FDG scanning as well as noncontrast CT scanning.

2. **PET scanning with ^{18}F-FDA** PET scanning may also be performed using radioisotope-tagged dopamine: 6-[^{18}F] fluorodopamine (^{18}F-FDA-PET). This scan is more specific for PGL and metastatic PHEO than is ^{18}F-FDG-PET/CT, because dopamine is a substrate for the norepinephrine transporter in tumor tissue. The sensitivity of ^{18}F-FDA-PET is 78% for nonmetastatic PHEO/PGL and 76% for metastatic PHEO/PGL. It is not widely available, but is generally more sensitive than ^{123}I-MIBG scanning, particularly in patients who harbor VHL or MEN 2 germline mutations. It is particularly useful for detecting other neuroendocrine tumors and PHEO/PGL metastases that are not visualized by scanning with ^{123}I-MIBG or ^{18}F-FDG and for tumors and metastases in patients with VHL disease.

E. Somatostatin receptor (SSTR) imaging SSTRs   are cell-surface G protein–coupled receptors, and there are five subtypes. About 70% of PHEOs express SSTRs, particularly types 2 and 4. SSTR imaging detects some metastases not visible on MIBG scanning, and vice versa. SSTR imaging has been reported to detect a cardiac PGL that was not visible on MIBG scanning.

111In-octreotide: Octreotide is a stable 8 amino acid analog of somatostatin with a high affinity for type 2 receptors. For imaging, octreotide is coupled with 111In-diaminetriaminepentacetate (DTPA). 111In has a half-life of 2.8 days and γ emissions of 173 keV and 247 keV. Scanning with 111In-labeled octreotide, known as SSTR imaging (SRI), has a sensitivity of only 25% for adrenal PHEOs and juxtarenal PGLs. This low sensitivity is due to high uptake of 111In-octreotide by the kidneys as well as renal excretion. However, 111**In-octreotide** detects 87% of PHEO metastases and is also a sensitive technique for detecting PGLs of the head and neck (chemodectomas). 111In-octreotide has physiologic uptake in the kidneys, thyroid, pituitary, salivary glands, gallbladder, bowels, spleen, liver, and mammary glands. Infection and recent surgery sites can also have misleading uptake of 111In-octreotide.

68Ga-DOTATATE-PET: This scans employs a somatostatin congener with increased affinity for SSTRs. It is tagged with 68Ga, a positron emitter, to allow PET scanning; when fused with a simultaneous CT scan; this provides exceptional sensitivity and clarity in the fusion scan. This scan is proving to be superior to all imaging modalities, particular for detecting PHEO/PGL metastases in patients with SDHB germline mutations.

F. Ultrasound imaging   On transabdominal ultrasound, an adrenal PHEO typically appears as a well-defined mass. Large PHEOs tend to develop internal hemorrhagic necrotic cysts, making the tumor appear heterogenous. Transabdominal ultrasound is most sensitive in slender individuals, in whom 85% of adrenal PHEOs can be visualized. However, ultrasound lacks specificity, such that a PHEO is not distinguishable from an adrenal adenoma or a mass in the superior pole of a kidney. Likewise, a left adrenal PHEO may be mistaken for a pancreatic tail tumor, and a right adrenal PHEO may be mistaken for a hepatic mass. Ultrasound may be used for initial imaging for PHEO in pregnant women, infants, and young children, although MRI is superior. Ultrasound is also useful for imaging and surveillance of neck PGLs. Ultrasound examinations have also been performed endoscopically, from the stomach and duodenum, with a longitudinal sector array, yielding sensitive detection of small adrenal PHEOs, lymphangitic metastases, and local recurrences. For pelvic and bladder PGLs, pelvic transvaginal ultrasound is very helpful for tumor localization and surveillance.

INCIDENTALLY DISCOVERED ADRENAL MASSES

Clinically unapparent adrenal nodules are discovered incidentally on about 4% of all abdominal CT or MRI scans that are performed for unrelated reasons (see Computed Tomography, discussed earlier.) Such nodules are known as **adrenal incidentalomas**. The incidence of adrenal nodules increases with age, being about 3% in middle-age and rising to about 10% in the sixth and seventh decades of life. Most such nodules are small, benign adrenal adenomas, with densities below 10 HU on noncontrast CT. However, PHEOs can produce nonspecific symptoms of abdominal pain, nausea, and weight loss, for which CT scans may be performed.

PHEOs account for about 4% of incidentally discovered adrenal masses. In the United States nearly half of the PHEOs diagnosed during life are detected incidentally on an abdominal or chest CT scan performed for other reasons. Therefore, most patients with adrenal nodules, even normotensives, should be screened for PHEO with plasma-fractionated free metanephrines. However, such testing is not mandatory for normotensives whose adrenal nodule has all of the following characteristics: (1) density on unenhanced CT less than or equal to 10 Hounsfeld Units (HU); (2) size less than or equal to 3 cm; (3) morphology not suspicious. A problem arises with metanephrines that are marginally elevated. Such patients with metanephrines that are one to two times higher than the reference range's ULN have about a 30% chance of harboring a PHEO. Those with metanephrines more than twice the ULN have a very high risk of having a PHEO. For apparently nonfunctioning adrenal nodules, it is generally reasonable to observe those that are under 3 cm in maximum diameter. Nonfunctioning nodules that are 3 to 5 cm in maximum diameter require especially close surveillance. Adrenal nodules that are over 5 cm in diameter are generally resected, except for obvious myelolipomas. When a PHEO has been ruled out, patients with an adrenal nodule may be screened for hyperaldosteronism and Cushing syndrome. Of note, hypertensive crisis has been described in patients with PHEO/PGL undergoing the high-dose dexamethasone suppression (DST) test, but not the low-dose DST.

Adrenal Percutaneous Fine-Needle Aspiration (FNA) Cytology

Most PHEO/PGLs can be readily diagnosed on the basis of their clinical, biochemical, and radiologic presentation. FNA cytology is not usually required for the diagnosis of a PHEO/PGL. However, some PHEO/PGLs are discovered incidentally on abdominal CT or ultrasound and may be clinically or biochemically silent. Although there may be a temptation to biopsy such masses, patients with a suspicious adrenal or retroperitoneal mass require testing for PHEO/PGL before any biopsy. There is a 70% risk of complications after percutaneous FNA of PHEO/PGLs. Such complications include: increased difficulty in the tumor's resection (41%), severe hypertension (15%), hematoma (30%), severe pain (25%), and incorrect or inadequate biopsy (25%). FNA cytology can be misinterpreted as a different primary malignancy or a metastasis from another malignancy; this potential for confusion is due to the fact that PHEO/PGLs are rare tumors and have pleomorphic and hyperchromic nuclei. Large left-sided PHEOs have been misdiagnosed as carcinoma of the tail of the pancreas based on CT scanning and FNA cytology. Percutaneous FNA can also disrupt the PHEO capsule and cause seeding of the tumor within the peritoneum.

Medical Management of Patients with Pheochromocytoma and Paraganglioma

Patients need to be treated with oral antihypertensives and stabilized hemodynamically prior to surgery. Patients receiving increasing doses of antihypertensive medications should have daily

measurements of blood pressure and pulse rate in the lying, sitting, and standing positions. Additionally, patients are taught to determine their own blood pressure and pulse rate regularly and during any paroxysmal symptoms. Most clinicians gradually increase antihypertensive medications over 2 or more weeks. However, prolonged preoperative preparation has not proven more effective for preventing intraoperative hypertension than shorter preparation periods. Some hypertensive patients have been admitted emergently for hypertension control and hydration, stabilized, and operated on successfully with intravenous infusion of a vasodilator drug (eg, nicardipine, nitroprusside, nitroglycerin; discussed later). Ideally, the blood pressure should be reduced to an average of 130/85 mm Hg (sitting) or less prior to surgery, while avoiding symptomatic orthostasis. Normotensive patients with PHEO/PGL who have documented increased plasma catecholamine levels are usually prepared for surgery with careful titration of α-blockade and/or calcium channel blockade. However, some normotensive patients, particularly those with PGLs, have truly nonsecreting tumors with completely normal plasma levels of catecholamines and metanephrines; such patients may not require any preoperative antihypertensive preparation, unless they have a clinical history of intermittent hypertension. Similarly, normotensive patients with PGLs secreting only dopamine do not require preoperative antihypertensive medication.

The type of preoperative medical preparation varies by region. In the USA, α-blockade is predominantly used. In Europe, about half the patients receive either nonselective α-adrenergic blockade (phenoxybenzamine) or selective α_1-adrenergic blockade (prazosin); about half receive calcium channel blockade (CCB), usually with nicardipine. Perioperative morbidity is the same with either type of medical preparation. The choice between preoperative α-blockade versus CCB (or a combination of both) is usually determined by local protocols and consultation between the endocrinologist, surgeon, and anesthesiologist and by the individual patient's reaction to the medication.

A. Alpha-adrenergic blockers   Alpha-adrenergic blockers have historically been used for most patients with PHEO in preparation for surgical resection. **Phenoxybenzamine** (10-mg capsules) is an oral nonselective α-blocker that is the most commonly used α-blocking agent. It has a long half-life of about 24 hours. Patients with mild hypertension may be given phenoxybenzamine at a starting dosage of 10 mg once daily, while those with more severe hypertension may receive a starting dosage of 10 mg twice daily. The dose of phenoxybenzamine may be increased by 10 mg every 2 days until the blood pressure falls to an average of 130/85 mm Hg sitting or until symptomatic orthostatic hypotension occurs. Patients who have been normotensive between paroxysms are particularly prone to develop hypotension with phenoxybenzamine. Phenoxybenzamine does not block the synthesis of catecholamines; in fact, the synthesis of catecholamines and metanephrines tends to increase during α-blockade. Therapy with phenoxybenzamine increases the heart rate but decreases the frequency of ventricular arrhythmias. Patients are encouraged to hydrate themselves well. Patients must be monitored daily for symptomatic orthostatic hypotension. Certain adverse effects are

common, including dry mouth, headache, diplopia, inhibition of ejaculation, and nasal congestion. Nasal decongestants should not be used if urinary catecholamine determinations or ^{123}I-MIBG scanning is planned, but antihistamines are acceptable. Phenoxybenzamine is not usually well tolerated as chronic therapy for hypertension in patients with unresectable or metastatic PHEO, and such patients are better treated with calcium channel blockers (discussed later), sometimes together with low-dose α-blockade.

Phenoxybenzamine crosses the placenta and accumulates to levels that are 60% higher in the fetus than in the maternal circulation; this can cause hypotension and respiratory depression in the newborn for several days following birth. Most patients require 30 to 60 mg/d, but the dosage is sometimes escalated to as high as 200 mg/d. However, phenoxybenzamine has a long half-life and excessive α-blockade with phenoxybenzamine can cause prolonged postoperative hypotension. Furthermore, excessive α-blockade may deny a critical surgical indicator (ie, a drop in blood pressure after complete resection of the tumor and aggravation of hypertension during palpation of the abdomen in case of multiple tumors or metastases).

Selective α_1-adrenergic blockers cause relatively less reflex tachycardia and orthostasis, compared to phenoxybenzamine. They also tend to cause less ejaculatory disturbance in men, which is a consideration with long-term use. **Doxazosin** is selective α_1-adrenergic blocker with a half-life of about 22 hours. It is effective in the medical management of PHEO/PGL when given orally in doses of 2 to 16 mg daily. In one series, there was no difference in hemodynamic instability during surgery in patients pretreated with doxazosin versus phenoxybenzamine. **Terazosin** is a selective α_1-adrenergic blocker with a half-life of about 12 hours. Although its usual indication is for benign prostatic hypertrophy, it has been used as an antihypertensive for patients with PHEO/PGL; the starting dose is 1 mg daily with subsequent doses being given twice daily and titrated upward every 2 days to a dose where hypertension is controlled or to a maximum of 10 mg every 12 hours. **Prazosin** is a short-acting selective α_1-adrenergic blocker. Due to its shorter half-life of only 3 hours, prazosin must be given every 8 to 12 hours. The starting dose of prazosin is 0.5 mg every 12 hours, increasing up to 10 mg twice or three times daily, if necessary. Patients treated with prazosin have experienced less postoperative hypotension, due to the drug's short half-life. Prazosin is particularly useful as a perioperative medication for pregnant women with PHEO/PGL who undergo a cesarean section (discussed later). **Urapidil** is an intravenous α_1-adrenergic blocker that has been used for in-hospital control of hypertension prior to surgical resection of PHEOs/PGLs. Urapidil has been successfully administered at constant intravenous infusion rates of 5 mg/h for the first day, increasing (if necessary) to 10 mg/h on the second day, and to 15 mg/h on the third day. In one study, urapidil therapy achieved preoperative hypertension control more rapidly (compared to phenozybenzamine), without any change in clinical outcomes.

Labetalol and carvedilol are combined selective α_1-adrenergic blocker *and* nonselective β-adrenergic blocking drugs with half-lives of about 6.5 and 8 hours, respectively. They should not be

used as the initial medication for treating hypertension in patients with PHEO/PGL, due to their unpredictable initial effects, sometimes causing an initial *increase* in blood pressure due to their nonselective β-blockade. Labetalol also can cause misleading elevations in catecholamine determinations in some assays, and inhibits MIBG uptake into PHEO/PGLs. However, labetalol has proven useful for some patients with metastatic or unresectable PHEO/PGL, especially after other medical therapy has failed. Labetalol is initiated at doses of 100 mg orally every 12 hours, titrating the dose upward every 2 days until the patient is normotensive or to a maximum of 1200 mg every 12 hours.

B. Calcium-channel blockers (CCBs)

Dihydropyridine CCBs are usually excellent antihypertensive agents for patients with PHEO/PGLs. However, they tend to reduce cardiac output and are not preferred for patients with congestive heart failure. Otherwise, CCBs are particularly useful for patients in whom α-blockers have caused adverse reactions, such as orthostatic hypotension or excessive fatigue. CCBs may be added to α-blockers to prepare patients for surgery. Patients who are normotensive between paroxysms are less likely to become hypotensive with CCBs compared to α-blockers. Patients with angina from coronary vasospasm are also best treated with CCBs. Perioperative fluid requirements have been lower among patients who were pretreated with CCBs instead of α-blockers. In a French series, 70 patients with PHEO were successfully prepared for surgery using oral CCBs (usually nicardipine). Nicardipine may be given in doses of 20 to 40 mg orally every 8 hours; nicardipine is also available as a sustained-release preparation that may be given in doses of 30 to 60 mg orally every 12 hours. Patients tend to tolerate CCBs better than α-blockers, such that CCBs are usually preferred for patients with PHEO/PGL who require long-term therapy due to recurrent or metastatic disease. However, some patients prefer α-blockade.

Nifedipine is a calcium channel blocker that is administered as an extended-release (ER) preparation in doses of 30 to 90 mg orally once or twice daily. The half-life of the ER preparation is 7 hours, so twice-daily dosing is preferred. For hypertensive paroxysms (systolic BP persistently above 170 mm Hg), nifedipine 10 mg (chewed pierced capsule) is usually a fast and effective treatment. Chewed nifedipine is generally safe for use by patients with PHEO, who may self-administer the drug at home during paroxysms but only with close blood pressure monitoring. In one small study, nifedipine therapy appeared to improve the uptake of MIBG into PHEOs in 4 of 8 patients at scanning doses. Nifedipine might possibly reduce the growth of PHEO/PGL and metastases, because *in vitro* nifedipine added to cultured PHEO cells reduces their mitotic index and proliferation. However, the effect of nifedipine on PHEO/PGL tumor growth has not been studied in patients.

Amlodipine may also be used for patients with PHEO/PGL in doses of 10 to 20 mg daily, but has a long half-life of 30 to 50 hours. There is little experience with other non-dihydropyridine CCBs (nimodipine, nisoldipine, isradipine, felodipine) for treating patients with PHEO/PGL.

Intravenous dihydropyridine CCBs are useful for hypertensive crisis, particularly during surgery. Nicardipine and clevidipine are both available as intravenous preparations (see Surgical Management of Pheochromocytoma and Paraganglioma, discussed later.)

Non-dihydropyridine CCBs are reported to be less effective for patients with PHEO. Verapamil has been used in a sustained-release preparation, but has been reported to be associated with postoperative pulmonary edema after PHEO resection. Diltiazem provides inadequate intraoperative blood pressure control.

C. Beta-adrenergic blockers

These agents are generally not prescribed for patients with PHEO/PGL tumors until treatment has been started with antihypertensive medications such as α-adrenergic blockers or CCBs. Beta-adrenergic blockade should then be used for treatment of β-adrenergic symptoms such as flushing, pounding heart, or tachycardia. It is important to institute α-blockade first, because blocking vasodilating $β_2$ receptors without also blocking vasoconstricting $α_1$ receptors can lead to hypertensive crisis if serum norepinephrine levels are high.

Nonselective β-blockers block both $β_1$- and $β_2$-adrenergic receptors. The inhibition of vasodilating arterial $β_2$ receptors causes unopposed vasoconstrictive α-adrenergic stimulation that aggravates hypertension. Therefore, nonselective β-blockers should ordinarily not be administered to patients with PHEO/PGL. Nonselective β-blockers include: nadolol, pindolol, propranolol, and timolol. Labetalol and carvedilol are different nonselective β-blockers that additionally block $α_1$ receptors.

Selective β-blockers specifically block $β_1$-adrenergic receptors at low doses. This leads to a rather selective reduction in heart rate without unopposed α-receptor dependent hypertension. However, at higher doses, these β-blockers also block $β_2$-adrenergic receptors and can cause a paradoxical worsening of hypertension. Selective β-blockers include: metoprolol succinate ER, atenolol, betaxolol, bisoprolol, nebivolol, and esmolol. Nebivolol has vasodilating activity through enhancement of nitric oxide release. Esmolol is the preferred intravenous preparation.

D. Angiotensin-converting enzyme inhibitors

ACE inhibitors have successfully treated hypertension in patients with PHEO/PGL, but not as the sole agent. Angiotensin receptor blockers have also been successfully added to multidrug antihypertensive therapy. Catecholamines stimulate renin production. In turn, renin stimulates the production of angiotensin I, which is converted by ACE to angiotensin II; this can be blocked by ACE inhibitors. Furthermore, PHEOs have been demonstrated to have angiotensin II-binding sites. ACE inhibitors are contraindicated in pregnancy (discussed later).

E. Metyrosine (α-methyl-para-tyrosine)

Metyrosine inhibits the enzyme tyrosine hydroxylase, which catalyzes the first reaction in catecholamine biosynthesis. Because of its potential side effects, it is usually used only to treat hypertension in patients with metastatic PHEO. Metyrosine is administered orally as 250-mg capsules, beginning with one every 6 hours; the dose is titrated

upward every 3 to 4 days according to blood pressure response and side effects. Most patients can tolerate 2 g/d, but higher doses usually cause side effects. The maximum dosage is 4 g/d. Catecholamine excretion is usually reduced by 35% to 80%. Preoperative treatment with metyrosine tends to reduce intraoperative hypertension and arrhythmias; however, postoperative hypotension is likely to be more severe for several days. Side effects of metyrosine include sedation, psychiatric disturbance, extrapyramidal symptoms, and potentiation of sedatives and phenothiazines. Crystalluria and urolithiasis can occur, so adequate hydration is mandatory. Metyrosine does not inhibit MIBG uptake by the tumor, allowing concurrent ^{123}I-MIBG scanning or treatment with high-dose ^{131}I-MIBG.

F. Octreotide   Octreotide has not been formally studied or approved for use in patients with PHEO/PGL. However, octreotide (100 μg subcutaneously three times daily) has been reported to reduce hypertensive episodes and catecholamine excretion in a man with PHEO whose hypertensive paroxysms were uncontrolled using other means. Octreotide therapy has been observed to reduce bone pain in a woman with a malignant PGL whose skeletal metastases were avid for ^{111}In-labeled octreotide. Octreotide therapy is usually begun at a dose of 50 to 100 μg injected subcutaneously every 8 hours. Side effects are common and may include nausea, vomiting, abdominal pain, and dizziness. If the drug is tolerated, the dose can be titrated upward to a maximum of 1500 μg daily. Octreotide LAR can be given as subcutaneous injections in doses of 10 to 40 mg every 4 weeks.

G. Activities to avoid   Vigorous exercise, particularly involving bending or heavy lifting, can aggravate hypertension in some individuals with PHEO/PGL, so mild exertion or rest is best during preparation for surgery. Emotional stress can provoke hypertensive attacks, so arguments and stressful situations are best avoided.

H. Foods to avoid   Intravenous tyramine causes hypertension in most patients with PHEO/PGL. While dietary tyramine has not been studied in patients with secretory PHEO/PGL, it is known that dietary tyramine can provoke hypertensive crisis in patients taking MAO inhibitor antidepressants. Therefore, although strict dietary precautions are not required, it is reasonable for patients with a known secretory PHEO/PGL to avoid consuming large amounts of foods with high tyramine content during preparation for surgery such as red wine, tap beers, aged dairy products, aged meats, fermented or pickled fish, liver, protein extracts, overripe fruit, soybeans, tofu, fava bean pods, bean pastes, brewer's yeast pills, marmite, and vegemite.

Surgical Management of Pheochromocytoma and Paraganglioma

A. Perioperative preparation   Prior to surgery, patients should be reasonably normotensive on medication, with an average blood pressure of 130/85 mm Hg or less, without symptomatic orthostasis (discussed earlier). They should also be well hydrated. It is ideal for patients to be admitted for administration of intravenous fluids at least 1 day prior to surgery; however, this is rarely done due to cost considerations. Patients may predonate blood for autologous transfusion. The transfusion of two units of blood within 12 hours before surgery reduces the risk of postoperative hypotension. Blood pressure must be monitored continuously during surgery. This requires placement of an arterial line, preferably in a large artery that is not prone to spasm (eg, femoral artery). A central venous pressure line helps determine the volume of fluid replacement. For certain high-risk patients with congestive heart failure or coronary artery disease, a pulmonary artery (Swan-Ganz) line may be inserted preoperatively to further optimize fluid replacement. Constant electrocardiographic monitoring is mandatory. Severe hypertension can occur—even in fully blocked patients—during bladder catheterization, intubation, or surgical incision. During laparoscopic surgery, catecholamine release is typically stimulated by pneumoperitoneum and by tumor manipulation. However, laparoscopic procedures cause less fluctuation of catecholamine levels and blood pressure than do open surgeries. All antihypertensive medications (discussed later) that might be required should be available and should be in the operating room well in advance.

Arterial embolization of the tumor can be performed immediately prior to surgery. This may be beneficial to the surgeon, particularly for the open resection of very large PGLs that tend to be extremely vascular and difficult to resect, due to blood loss and venous oozing. Embolization of neck PGLs is an established intervention. Secretory PGLs can also be embolized preoperatively. An experienced interventional radiologist must perform such embolizations with full monitoring and an anesthesia team in attendance.

The major problem during surgery for PHEO/PGL is hemodynamic instability. Serious blood pressure variations are more common in patients whose blood pressure has not been adequately controlled preoperatively. Intraoperative hemodynamic instability is also more common in patients with higher plasma norepinephrine levels and larger tumor size.

B. Antihypertensive and antiarrhythmic drugs

1. **Calcium channel blockers** These are effective therapies for intraoperative hypertension. They can cause reflex tachycardia that can be controlled with intravenous β-blockers (eg, esmolol). Nicardipine is administered as an intravenous infusion, starting at a dose of 5 mg/h and increasing by 2.5 mg/h every 5 to 15 minutes up to 15 mg/h. Nicardipine was successfully used as the sole intraoperative vasodilating agent in one French series of 70 patients and in another series of 19 patients. Its half-life is 9 hours. Clevidipine has a shorter half-life of 1 to 15 minutes. It is available for intravenous infusion at a dose of 1 to 2 mg/h, doubling the dose after 90 seconds, then increasing in smaller increments at longer intervals (5-10 minutes) up to 4 to 6 mg/h or a maximum of 16 mg/h.

2. **Phentolamine** It is a parenteral α-adrenergic blocker that has a short half-life of 19 minutes. It can be given intravenously in bolus doses, starting with a trial dose of 2.5 mg, followed by 5 mg (children, 1-3 mg) and repeating in doses of 5 to 15 mg as

needed for blood pressure control. Phentolamine may also be given by intravenous infusion at a rate of 0.5 to 1 mg/min. Side effects include hypotension, tachycardia, cardiac arrhythmias, nasal stuffiness, nausea, and vomiting.

3. **Nitroprusside** It is given by intravenous infusion and is an effective drug for managing hypertensive episodes; advantages include widespread familiarity with its use and its short half-life of 2 minutes. Nitroprusside is initiated at 0.25 to 0.3 µg/kg/min and titrated for the desired effect. The maximum infusion rate is 10 µg/kg/min for 10 minutes. High infusion rates should not be given for prolonged periods; long-duration (over 6 hours) nitroprusside infusion rates above 2 µg/kg/min cause cyanide accumulation and toxicity. Cyanide toxicity causes metabolic acidosis and an increase in venous oxygen saturation (>90%) in severe cases. If cyanide toxicity is suspected, the nitroprusside infusion must be stopped or slowed. In the United States, there are two different cyanide antidote kits available. (1) The conventional kit contains amyl nitrite, sodium nitrite, and thiosulfate. Administer an amyl nitrite crushed ampule at the end of the endotracheal tube or under the patient's nose and give 10 mL 3% sodium nitrite intravenously. Also administer 25% sodium thiosulfate solution (50 mL) intravenously. (2) Additionally, hydroxocobalamin (Cyanokit) is available for intravenous administration at a dose of 5 g (children, 70 mg/kg). Coadministration of sodium thiosulfate (1 g/100 mg nitroprusside) prevents cyanide accumulation. Nitroprusside must be administered cautiously, since it can cause precipitous and profound hypotension, resulting in irreversible ischemic injuries.

4. **Nitroglycerin** It is given by intravenous infusion and is effective therapy for perioperative hypertension. Nitroglycerin infusions are initiated at 5 µg/min and increased by 5 µg/min every 3 to 5 minutes until target blood pressure is achieved or a dose of 20 µg/min is reached; if the response has been insufficient, the dose can be increased by 10 to 20 µg/min every 5 minutes to a maximum of 100 µg/min. Because nitroglycerin adheres to polyvinyl chloride tubing, non-PVC infusion sets must be used. Nitroglycerin infusions can cause headache and hypotension. Methemoglobinemia has occurred during prolonged high-dose infusions and is manifested by cyanosis in the presence of a normal arterial pO_2. The therapy for methemoglobinemia consists of immediately stopping the nitroglycerin and giving methylene blue, 1 to 2 mg/kg intravenously.

5. **Beta-blockers** Atrial tachyarrhythmias may be treated with intravenous atenolol in 1-mg boluses or by constant intravenous infusion of esmolol, a short-acting β-blocker. Esmolol is given as an initial dose of 500 µg/kg intravenously over 1 minute, then continued at 50 µg/kg/min. If required, the infusion rate may be increased by 50 µg/kg/min every 4 minutes.

6. **Magnesium sulfate** It is useful for suppressing ventricular arrhythmias and also for managing hypertension during resection of a PHEO/PGL, particularly during pregnancy. Magnesium sulfate may be given as a 4-g intravenous bolus, followed by 3 to 20 mg/min IV infusion.

7. **Lidocaine** It may be used to treat cardiac ventricular arrhythmias. In adults, lidocaine is administered with a loading dose of 150 to 200 mg over about 15 minutes, or as a series of smaller boluses. This is followed by a maintenance infusion of 2 to 4 mg/min in order to achieve a therapeutic plasma level of 2 to 6 µg/mL.

8. **Drugs to avoid** *Atropine* should not be used as preoperative medication for patients with PHEOs because it can precipitate arrhythmias and severe hypertension. Metoclopramide, glucagon, and glucocorticoids can also precipitate a hypertensive crisis. *MAO inhibitor antidepressants* can provoke hypertensive crisis by blocking the metabolism of catecholamines. Other medications that can elicit hypertensive crisis include decongestants (eg, pseudoephedrine), epinephrine, amphetamines and amphetamine derivatives, and cocaine. *Labetalol* is not recommended for preoperative or intraoperative management of PHEOs because it may aggravate post-resection hypotension. It can also paradoxically aggravate hypertension early in the course of treatment, because its β-blocking effect may occur initially, allowing a brief period of unopposed α-receptor stimulation. Labetalol also inhibits MIBG uptake and causes misleading elevations in catecholamine determinations in certain assays. *Diazoxide* has been ineffective against hypertension caused by PHEO/PGL.

C. Anesthesia Anesthetic agents such as intravenous propofol, enflurane, isoflurane, sufentanil, alfentanil, and nitrous oxide appear to be safe and effective. Atropine should not be used. Muscle relaxants with the least hypertensive effect should be employed (eg, vecuronium). Intraoperative hypertension can be managed by increasing the depth of anesthesia and by intravenous vasodilators for blood pressures over 160/90 mm Hg. Serum catecholamine levels drop sharply after adrenal vein ligation and profound hypotension can occur suddenly after resection of a PHEO. Therefore, it is prudent to stop the vasodilator infusion just prior to adrenal vein ligation.

D. Operative management Perioperative mortality is about 2.4% overall, but morbidity rates of up to 24% have been reported. Surgical complications do occur and include splenectomy, which is more common with open abdominal exploration than with laparoscopic surgery. Reported surgical complication rates have been higher in patients with severe hypertension and in patients having reoperations. Surgical morbidity and mortality risks can be minimized by adequate preoperative preparation, accurate tumor localization, and meticulous intraoperative care.

1. **Laparoscopy** Most PHEOs can be resected laparoscopically, which has become the procedure of choice for removing most adrenal neoplasms that are under 6 cm in diameter. Adrenal laparoscopic surgery is usually performed through four subcostal ports of 10 to 12 mm. Laparoscopic surgery is widely used now that preoperative localization of the tumor is possible. However, tumors that are invasive or over 6 cm in diameter are more difficult to resect laparoscopically and may require open surgery. For larger PHEOs/PGLs, a lateral laparoscopic approach can be used, because it affords greater opportunity to explore the abdomen and inspect the liver for metastases. For patients with a small adrenal PHEO and for those who have had prior abdominal surgery, a posterior laparoscopic approach may be preferred.

The laparoscope allows unsurpassed magnified views of the PHEO and its vasculature. PHEOs are bagged to reduce the risk of fragmentation and spread of tumor cells within the peritoneum or at the port site. Larger tumors can be removed through laparoscopic incisions that can be widened for the surgeon's hand (laparoscopic-assisted adrenalectomy). With laparoscopic surgery, hypotensive episodes are less frequent and

less severe. Laparoscopic adrenalectomy has other advantages compared with open adrenalectomy: less postoperative pain, faster return to oral foods, and shorter hospital stays (median 3 days vs 7 days for the open approach). This approach is the least invasive for the patient, who can usually begin eating and ambulating the next day. The laparoscopic approach may be used during pregnancy. The technique has also been used successfully for certain extra-adrenal PGLs. Surgical-related mortality has been reduced to 3% at referral centers.

2. **Needlescopic adrenalectomy** This procedure uses three sub-costal ports of 3 mm, with a larger umbilical port for tumor removal. In one series of 15 patients, this technique reduced surgical times and recovery time compared with the standard laparoscopic approach. However, more extensive experience with this technique is required.

3. **Adrenal cortex-sparing surgery** All patients undergoing bilateral total adrenalectomies require life-long glucocorticoid and mineralocorticoid hormone replacement. To avoid adrenal insufficiency, patients with bilateral PHEOs and those with unilateral PHEOs but with highly penetrant familial PHEO/PGL syndromes (especially MEN 2 and VHL) who are at high risk for developing a contralateral PHEO, have had successful laparoscopic partial adrenalectomy during resection of small adrenal PHEOs, sparing some of the adrenal cortex. Such adrenal-sparing surgery is difficult and risks a local recurrence of the PHEO, especially in patients with MEN2.

4. **Open laparotomy** Open laparotomy is indicated for patients with particularly large PHEOs or PGLs, or for those with intra-abdominal metastases that require debulking. Large vascular PGLs can be considered for preoperative arterial embolization, but its efficacy is uncertain. An open anterior midline or subcostal approach usually yields adequate exposure. For patients with PGLs of the urinary bladder, a partial cystectomy can sometimes be curative. Other patients with larger bladder PGLs require a total cystectomy and construction of a diverting ureteroenterostomy if the tumor has not been fully resected. For patients with a curative total cystectomy, construction of a new ileal neobladder is possible.

E. Intravenous dextrose   Immediately following removal of a secretory PHEO/PGL, intravenous 5% dextrose should be infused at a constant rate of about 100 mL/h to prevent postoperative hypoglycemia that is otherwise frequently encountered.

F. Therapy for shock occurring after PHEO/PGL resection   Severe shock and cardiovascular collapse can occur immediately following ligation of the adrenal vein during resection of a PHEO, particularly in patients having norepinephrine-secreting tumors. Such hypotension may be due to desensitization of α_1-adrenergic receptors, persistence of antihypertensives, and low plasma volume. Preoperative preparation with CCBs or α-blockade plus intravenous hydration or blood transfusions reduces the risk of shock. Intravenous antihypertensives are held just prior to ligation of the adrenal vein. Treatment of shock consists of large volumes of intravenous saline or colloid. Intravenous norepinephrine is sometimes required in very high doses. In cases of post-resection catecholamine-resistant hypotension, intravenous low dose vasopressin administration may be efficacious, given as boluses of 0.08 U, followed by an intravenous continuous vasopressin infusion of 1.6 U/h.

Pregnancy and PHEO/PGL

PHEO/PGLs occur in about 1 in every 50,000 pregnancies and are often unrecognized antepartum. Such unsuspected PHEO/PGLs result in a maternal mortality of 40% and a fetal mortality of 56%. However, if the diagnosis is made antepartum, the mortality is much lower. Hypertension is often misdiagnosed as eclampsia or preeclampsia. Hypertensive crisis usually occurs during labor and can be associated with cardiac dysrhythmias or ARDS. Although maternal catecholamines do not cross the placenta, maternal hypertensive crisis is very dangerous for the fetus, causing uteroplacental insufficiency and fetal death.

Pregnant hypertensive patients are often treated with methyldopa that can cause false-positive testing for catecholamines by older fluorometric methodologies, but not by HPLC. Methyldopa does not cause interference with plasma or urine metanephrine measurements performed with MS/MS. Labetalol is often used to treat pregnant hypertensive women, but its use with PHEO/PGL is discouraged (discussed earlier).

During pregnancy, the localization of a PHEO in pregnancy is best done with MRI. As soon as a PHEO is diagnosed, α-blockade is commenced. **Phenoxybenzamine** is usually used. However, phenoxybenzamine crosses the placenta and accumulates in the fetus. After 26 days of maternal phenoxybenzamine therapy, cord blood levels in the newborn are 60% higher than the mother's serum levels. Therefore, some perinatal depression and hypotension may occur in newborns of mothers receiving phenoxybenzamine. Prazosin has an advantage for maternal treatment near term, since it is a short-acting selective α_1-adrenergic blocker and causes less newborn hypotension, compared to longer acting α-adrenergic blockers. However, chronic use of prazosin increases the risk of fetal demise. The starting dose of prazosin is 0.5 mg/d orally, increasing up to 10 mg orally twice daily if necessary.

There has been little experience with the use of CCBs to treat hypertension in pregnant women with PHEO. However, CCBs can be safely added to α-blockers during pregnancy and are not teratogenic in the first trimester. Therefore, nifedipine or nicardipine may be used to supplement or replace α-blockade as needed.

If possible, β-blockade should not be used at all during pregnancy. Propranolol crosses the placenta and can cause intrauterine growth restriction. Newborns of mothers taking propranolol at delivery exhibit bradycardia, respiratory depression, and hypoglycemia. Therefore, β-blockers should ideally be discontinued 48 hours prior to expected delivery. During delivery, serious atrial tachyarrhythmias should be controlled by a short infusion of esmolol, a β-blocker with a very short half-life.

Angiotensin converting enzyme (ACE) inhibitors should not be used during pregnancy, since their use in the second and third trimesters has been associated with fetal malformations, including skull hypoplasia, renal failure, limb and craniofacial deformation, lung hypoplasia, intrauterine growth retardation, patent ductus arteriosus, and death.

During the first 6 months of pregnancy, it is often possible to treat a woman with α-blockade, followed by laparoscopic resection of the tumor. Although the fetus usually survives, spontaneous abortion is common, despite a successful resection of the

tumor. If a PHEO is not discovered until the last trimester, treatment consists of α-blockade followed by elective cesarean delivery as early as feasible. Intravenous magnesium sulfate is a useful antihypertensive during surgery. Intravenous CCBs may be used to treat patients with hypertensive crisis caused by PHEO in pregnancy. The tumor is resected after the cesarean delivery. In the presence of an active PHEO, vaginal delivery should never be allowed, since life-threatening hypertensive crisis will predictably occur.

Pheochromocytoma-Induced Life-Threatening Complications: Cardiomyopathy, ARDS, and Multisystem Crisis

Life-threatening complications can occur in patients with PHEO/PGL in an idiosyncratic fashion and are not necessarily associated with hypertension.

1. **Cardiomyopathy** and attendant serious cardiac dysrhythmias often occur in patients with PHEO/PGL. It is also known as "catecholamine cardiomyopathy." Heart failure and shock can occur and can sometimes present with reverse ventricular activity, also known as "Takatsubo cardiomyopathy." "Takatsubo" is a Japanese term for a traditional octopus fishing pot, which the heart may resemble on imaging studies. Affected patients can present with life-threatening ventricular arrhythmias, cardiac rupture, systemic embolism, or shock due to cardiac pump failure. Serum BNP can be very high. The etiology for cardiomyopathy in PHEO/PGL has been ascribed to catecholamines, although circulating cytokines may also be contributory. Acute cardiomyopathy is potentially reversible after the tumor is removed or the patient is stabilized. Therefore, every effort must be made to salvage these patients, who may require mechanical circulatory support (MCS).

2. **ARDS** occurs in patients with PHEO/PGL and is not associated with unusually severe hypertension or particularly high plasma catecholamine levels, compared to other PHEO/PGLs. ARDS is more apt to occur in patients with proteinuria that may serve as a marker for patients at risk. Proteinuria and ARDS itself may be due to tumoral secretion of certain cytokines in the setting of catecholamine excess. ARDS can occur spontaneously, after major surgery, and after therapy with high-activity ^{131}I-MIBG in patients with metastatic disease. ARDS may accompany cardiomyopathy and can be mistaken for congestive heart failure or pneumonia. Most patients require intubation and respirator-assisted breathing with high oxygen supplementation. Some patients may require extracorporeal membrane oxygenation (ECMO). ARDS may progress to multisystem crisis.

3. **Multisystem crisis** can have manifestations such as ARDS, renal failure, hepatic failure, rhabdomyolysis, cardiomyopathy, shock, and disseminated intravascular coagulation. Multisystem crisis can occur spontaneously or following the administration of glucagon or corticosteroids. It can also occur following unrelated surgery or therapy with ^{131}I-MIBG. The mortality rate is extremely high despite intensive care. It has been hypothesized that ARDS and multiorgan crisis may be due to increased circulating cytokines, in addition to catecholamines.

Pathology of PHEO and PGL

On histopathology, PHEOs are extremely vascular tumors with neurosecretory granules. Central necrosis is often present. The cells may be arranged in nests (zellballen pattern), anastomosing cords (trabecular pattern), or a combination of both. Cells vary in size with pleomorphic and eccentric nuclei that are often large and bizarre in appearance. The cells exhibit immunofluorescence staining for CgA and synaptophysin.

No single characteristic of PHEO or PGL can determine whether a given tumor is malignant. Therefore, the definition of malignancy is based upon whether metastases are present. Metastases must be distinguished from additional PGLs by their location (liver, lung, bone) where sympathetic paraganglia are rare. Metastases must also be distinguished from intraperitoneal seeding, a phenomenon known as pheochromocytomatosis. Metastases can vary in virulence from relatively indolent to extremely aggressive (see discussion on malignant PHEO later).

Overall, about 26% of these tumors arise in patients with identifiable germline mutations. Individuals with no known family history of these tumors have about a 17% risk of harboring a known germline mutation. About 18% of cases develop in children. The earlier a tumor presents, the more likely that individual harbors a germline mutation (see discussion on genetics of PHEO and PGL, earlier).

Adrenal PHEOs and PGLs appear very similar on microscopy. PGLs are often large and arise near the adrenal, making the distinction particularly difficult based upon preoperative localization scans. But it is important to distinguish these two tumors because PGLs are more likely to metastasize. One way to distinguish these tumors preoperatively is to evaluate their secretions. Tumors that secrete epinephrine (or its metabolite, metanephrine) are predictably adrenal PHEO. However, tumors that secrete strictly norepinephrine (or its metabolite normetanephrine) may be either PHEO or PGL. Intraoperatively, the surgeon may be able to identify the adrenal gland as separate from the tumor. On pathology, it is often possible to visualize the adrenal cortex lying in close proximity to the PHEO or even arising out of the adrenal medulla. An intact adrenal gland may be included in the surgical specimen, indicating that the tumor was a juxta-adrenal PGL. On microscopy, PGLs may have visible nerve ganglia.

Composite PHEOs are rare tumors that exhibit histopathologic features of both PHEO and neuroblastoma. Such tumors have generally not recurred and have not exhibited *N-myc* amplification, which distinguishes them from typical neuroblastomas. Therefore, composite PHEO are considered to be a histological variant of PHEOs.

Metastatic Pheochromocytoma and Paraganglioma (Table 11–18)

No PHEO or PGL should be labeled *benign*. Histopathology cannot reliably determine whether a given tumor has metastasized. Such metastases can be microscopic and indolent, eluding detection with the most sensitive scanning and biochemical screening. Therefore, it is best to think of these tumors as either having

TABLE 11–18 Distribution of metastases in 50 cases of metastatic pheochromocytoma and paraganglioma.

Region of Metastasis	Percentage
Bones	82
Abdomen (nodes, peritoneum, local recurrence)	56
Lung parenchyma	36
Liver	30
Pelvis (soft tissue)	20
Mediastinum and lung hila	18
Lymph nodes (extra-abdominal) especially supraclavicular and inguinal	14
Chest wall	12
Muscle	6
Brain	0

Data from Gonias S, Goldsby R, Matthay KK, et al. Phase II study of high-dose [131I] metaiodobenzylguanidine therapy for patients with metastatic pheochromocytoma and paraganglioma. *J Clin Oncol*. 2009 Sep 1;27(25):4162-4168.

Note: Some cases of abdominal recurrence are peritoneal seeding (pheochromocytomatosis).

detectable metastases (metastatic PHEO) or having *no detectable* metastases.

Location: Metastases are evident at the time of diagnosis in about 10% of patients with an adrenal PHEO. Another 5% to 10% are found to have metastatic disease or local recurrence within 20 years. Extra-adrenal PGLs commonly metastasize. In a Mayo Clinic series of PGLs, 15% had local invasion and 21% had detectable distant metastases at the time of initial surgery, with an overall 36% risk of local or distant metastasis.

Genetics: The risk for detectable metastases is high for PHEOs/PGLs that arise in patients harboring a *SDHB* germline mutation, with this mutation being found in about 40% of metastatic PHEOs/PGLs. PHEOs/PGLs that arise in patients with *NF-1* germline mutations have a 12% risk of metastasis. PHEOs that arise in patients with *VHL* or *RET* germline mutations have a low (<5%) risk of metastasis.

It is conceivable that *all* PHEOs and PGLs metastasize as single cells but that such cells will only grow if certain genes are down-regulated and other genes are up-regulated. This concept is supported by genome-wide expression profiling of PHEO/PGLs. These profiles have identified candidate genes that are differentially expressed in tumors with detectable metastases versus those with no detectable metastases. Gene expression arrays of PHEOs/PGLs with detectable metastases have demonstrated that 5 genes are differentially up-regulated in these tumors.

Research in malignant PHEO/PGL has been hampered by the absence of a viable human cell line. However, a highly malignant mouse PHEO model has been developed in which a gene expression array has demonstrated the up-regulation of 8 genes and the down-regulation of 38 genes.

Other predictors of malignancy: The risk of PHEO/PGL recurrence is higher if there is extensive local invasion. One surgical observation is that tumors that are "stickier" and more difficult to dissect from adjacent tissue are more likely to recur or metastasize. Tumors are also more likely to metastasize if they are larger than 5 cm diameter, although about 20% of metastatic primary tumors have been smaller than 5 cm diameter. Primary PHEOs/PGLs as small as 1 cm diameter have been documented to metastasize. Tumors are also more likely to metastasize if they contain high levels of Ki-67, a protein expressed in proliferating cells that can be detected by the monoclonal antibody MIB-1 and quantified as a high MIB-1 score. Metastatic PHEOs/PGLs also have increased activity of telomerase. Tumors with high *c-myc* gene expression are more likely to be malignant. In one series, 50% of patients with malignant PHEO were found to have high serum levels of NSE, but in none of 13 patients in another series with benign PHEOs.

The differential diagnosis for apparent metastases includes, multicentric PGLs, second PHEOs, false-positive scanning, and intraperitoneal seeding of tumor (pheochromocytomatosis). Patients with familial forms of PHEO/PGL may develop other tumors associated with their mutation: medullary thyroid cancer (*RET*), renal cell cancer (*VHL, SDHx*), pituitary adenoma, GIST, pulmonary chondroma (SDHx). Other non-PHEO/PGL malignancies also appear to be more common. A series of 110 Swedish patients with PHEO/PGL, treated surgically, showed an unexpectedly high relative risk for developing non-PHEO/PGL malignancies (RR ≅ 2.0). Therefore, additional tumors should not be presumed to be metastases unless they have uptake on [123I]-MIBG or [18]FDA-PET scanning. Suspicious lesions without such uptake should be considered for biopsy.

Sites of metastasis (Table 11–18): Metastases from PHEOs or PGLs typically involve bones (82%), liver (30%), lungs (36%), lymph nodes, the contralateral adrenal, and sometimes muscle. The bones most frequently involved include vertebrae, pelvis and ischium, clavicles, cranium, proximal femurs, and humeri. Tibia, radius, and ulna are occasionally involved. Patients with metastases and an *SDHB* mutation have a higher prevalence of metastases involving long bones compared to those without the mutation. PHEO/PGL metastases have a proclivity for the skull where they may form a dumbbell-type lesion, sometimes being palpable as a soft cyst-like bump; they also may grow inside the skull to impinge upon the brain. Prevertebral PGLs may destroy adjacent vertebrae, and spinal cord compression may occur. Most bone metastases affect the cortical bone; they may be indolent, but are often osteolytic and cause progressive bone destruction. Other bone metastases primarily involve trabecular bone and marrow; such metastases are usually visible on MRI but may be invisible on CT. Metastases to lymph nodes outside the abdomen are most frequent in the supraclavicular and inguinal regions. Metastases to muscle are usually indolent. These tumors have not been reported to metastasize to brain, although metastases to the cranium and skull base may impinge upon the brain, pituitary, and cranial nerves.

Metastases usually secrete norepinephrine and normetanephrine. Some metastases secrete predominantly dopamine. Metastases rarely secrete epinephrine or metanephrine, with the exception of some metastases from epinephrine-secreting adrenal PHEOs. About 20% of sympathetic PGLs and their metastases do not secrete catecholamines or metanephrines, but most continue to secrete CgA, which becomes a valuable tumor marker.

Assessing the growth rate of PHEO/PGL metastases: Treatment can be tailored to the tumor's rate of growth. Knowledge of the growth rate of a patient's metastases may be obtained through close biochemical surveillance and serial volumetric imaging with CT or MRI. Patients who are asymptomatic with a few indolent osseous metastases may elect to receive bisphosphonate therapy and delay life-threatening chemotherapy or radioisotope therapy, as long as they remain under close surveillance. Even patients with a symptomatic osteolytic metastasis may elect to receive targeted therapy rather than systemic treatment if their overall tumor burden is low.

Surveillance: Malignancy is determined only by the presence of metastases. MRI or CT scanning may not detect small metastases within their field and will certainly not visualize metastases outside their field. The various radiolabeled imaging scans can also be negative in the presence of metastases. Therefore, following the resection of an apparently benign PHEO or PGL, long-term surveillance is required, since metastases may not become clinically apparent for years or decades.

Biochemical surveillance: Patients should be followed with repeated determinations for the tumor markers that were highest prior to resection of their primary PHEO/PGL. Repeat postoperative testing should be postponed until at least 2 to 4 weeks postoperative, and the patient is fully recovered from surgery. Unfortunately, some patients with unsuspected PHEO/PGL have incomplete or no preoperative testing prior to the surgical resection of their primary tumor. Biochemical screening for recurrence or metastases in such patients is best done with plasma-fractionated free metanephrines and fractionated catecholamines (with dopamine), along with a fasting serum CgA. While this survey of tumor markers is sensitive for PHEO/PGL recurrence, false-positive testing occurs commonly. When a PHEO/PGL tumor marker is only marginally elevated, the test should be repeated. Interpretation of elevated tumor markers should be done cautiously with the knowledge that adrenal PHEOs usually secrete both norepinephrine and epinephrine, and their metastases may sometimes continue to produce epinephrine, but more commonly produce norepinephrine and its metabolite normetanephrine. Extra-adrenal PGLs ordinarily produce only norepinephrine and normetanephrine and sometimes only CgA; their metastases do likewise. When such tumor markers are consistently elevated, repeat scanning is certainly warranted.

Scan surveillance: It must be kept in mind that scan surveillance is useful to detect not only recurrent or metastatic PHEO/PGL, but also to detect other malignancies to which these patients are prone. Unfortunately, the long-term repetitive use of any CT or radionuclide scan delivers excessive cumulative radiation exposure to the patient, increasing their lifetime risk of additional non-PHEO/PGL malignancies. Therefore, a reasonable combination of modalities must be used, relying upon biochemical surveillance for secretory sympathetic PHEO/PGLs (discussed earlier). Neck ultrasound is useful for nonsecretory HN-PGLs, but will not detect a jugulotympanic PGL. MRI is useful for detecting recurrent or de novo HN-PGLs. MRI is also useful for general scanning, but scanning the entire chest, abdomen, and pelvis usually requires a prolonged time in the scanner; it can also be problematic to obtain insurance preauthorization and to schedule a full-body MRI. *Rapid full-body MRI* has been developed and is available in certain tertiary referral centers. Rapid full-body MRI reduces the time for each image cut from several seconds to milliseconds. However, it is a screening scan, rather than diagnostic procedure and can miss smaller tumors. Also, full body MRI does not typically cover the extremities and does not detect metastases to the extremities.

Radionuclide scanning is advantageous in that the entire body may be included in the scan. Some kind of radionuclide scanning is indicated for patients with biochemical or clinical evidence of recurrent or metastatic PHEO/PGL. However, no radionuclide scan is 100% sensitive. PHEO/PGL metastases have variable avidity for ^{123}I-MIBG. Even when the primary tumor is avid for ^{123}I-MIBG, some or all of the metastases may not be visible on ^{123}I-MIBG scanning. This low sensitivity (57%) is seen for metastases that are deficient in norepinephrine transporter (NET) expression. In a given individual, some metastases may be avid for ^{123}I-MIBG, while others show virtually no uptake. When patients have a recurrence or progression of metastases after ^{131}I-MIBG therapy, metastases with less MIBG-avid may emerge or progress in size. When ^{123}I-MIBG imaging is negative, metastases may be detected with MRI or CT scanning. Metabolically active PHEO/PGL metastases are usually visible with ^{18}F-FDG-PET scanning, which has the additional advantage of being a whole-body scan with a sensitivity of about 76% for metastases. ^{18}F-FDA-PET has a sensitivity of about 78% for metastases. Somatostatin receptor scintigraphy or ^{68}Ga-DOTATATE-PET, will often detect metastases that are not visible with other scans.

Treatment for Patients with Recurrent or Metastatic PHEO and PGL

Patients with recurrent or metastatic PHEO/PGL must have individualized therapy and surveillance. They are usually considered for surgery, with the goal of either surgical cure or at least debulking. Some small metastases can be quite indolent, such that certain asymptomatic patients with a low and indolent tumor burden may be followed closely without treatment or treatment with less toxic modalities, such as denosumab or zoledronic acid for osseous metastases, and directed external beam radiation therapy for larger osteolytic metastases. In one series of 90 patients with metastatic PHEO/PGL, 9% had progression-free survival at 5 years.

A. General considerations *In vitro* studies have found that cultured PHEO cells are protected from cytotoxic insult by amitriptyline and fluoxetine, possibly through up-regulation of superoxide dismutase. Therefore, patients being treated for metastatic PHEO/PGL should probably not take tricyclic antidepressants or selective serotonin reuptake inhibitors, although there have been no clinical studies of their effect on survival. A clinical study of kindreds with *SDHD* germline mutations noted that individuals with this mutation who lived at higher elevations developed more PGLs earlier than their affected relatives living at sea level (the Netherlands). Therefore, it is possible that residing at higher altitude may stimulate the development of PHEO/PGL among

patients harboring *SDHD* germline mutations and perhaps other cluster 1 germline mutations with underlying pseudohypoxia (see Table 11–8). There is concern that chronic recurrent hypoxia from sleep apnea can trigger catecholamine release and might also stimulate tumor growth in susceptible patients with cluster 1 germline mutations. Therefore, screening for sleep apnea should be considered for affected patients who snore or have other risks for sleep apnea.

B. Surgery   It is usually best to resect the primary tumor as well as large metastases. This is especially true of secretory tumors that are causing hypertension and other symptoms that can be life-threatening. Even lung metastases may be resected. However, the decision about whether to resect metastases is a difficult one and must be based upon very thorough staging of the patient's tumor. Of course, when resecting secretory metastases, preoperative preparation is mandatory, and hypertension must be adequately controlled.

C. Chemotherapy   Due to the rarity of PHEO/PGL, there have been no large-scale clinical trials comparing available chemotherapies. However, there have been small series of patients and case reports from which to derive recommendations. Although chemotherapy can achieve remissions, no durable long-term remissions have been reported, such that lifetime chemotherapy is required. Therefore, when selecting chemotherapy, it is important to select a regimen that the patient can tolerate long term with a reasonable quality of life. Temozolomide is usually tolerated best. However, for aggressive metastases, it is best to induce a remission with either a tyrosine kinase inhibitor or cyclophosphamide, vincristine, and dacarbazine (CVD) combined chemotherapy (discussed later).

Temozolomide (TMZ): TMZ carries the advantage of either oral or intravenous administration on a daily basis for just 5 days monthly. Patients usually tolerate TMZ better than other chemotherapeutic regimens for long-term tumor suppression. However, TMZ causes nausea, and patients require prophylactic ondansetron and other supportive therapy during their treatment week. TMZ has reasonable anti-tumor effect in patients with metastatic PHEO/PGL, particularly those with *SDHB* germline mutations. In one study of 15 patients with metastatic PHEO/PGL (10 with *SDHB* germline mutations), treated with a mean dose of TMZ of 172 mg/m^2 daily for 5 days monthly, partial responses (PR) were observed in 5 patients, stable disease (SD) in 7 patients, and progressive disease (PD) in 3 patients. All 5 patients experiencing a PR had an *SDHB* mutation. If most metastases remain stable, the patient may continue TMZ, while limited break-through metastases are treated surgically or with external beam radiation therapy.

Cyclophosphamide, vincristine, dacarbazine (CVD): Combined chemotherapy with CVD is usually given intravenously over 2 days and the cycles repeated every 3 to 4 weeks, indefinitely. After controlling symptoms of catecholamine excess, cycles of CVD are administered as follows: cyclophosphamide 750 mg/m^2, vincristine 1.4 mg/m^2, and dacarbazine 600 mg/m^2 on day 1, followed by dacarbazine 600 mg/m^2 on day 2. The doses are adjusted according to the response and toxicity in each patient.

In a meta-analysis of 50 patients with metastatic PHEO/PGL treated with CVD, the tumor-size response rates have been: CR 4%; PR 37%; SD 14%; PD 45%. Catecholamine response rates have been: CR 14%; PR 40%; SD 20%; PD 26%.

When CVD chemotherapy is stopped, the tumors usually recur. Many patients cannot tolerate such a long-term regimen, due to fatigue, cytopenias, neuropathy, and other adverse reactions. However, some patients tolerate it reasonably well and may experience a complete biochemical remission. Once in remission, CVD cycles are continued at increased intervals, but not stopped.

Sunitinib: Partial remissions in metastatic PHEO have been reported with sunitinib, a tyrosine kinase inhibitor. Sunitinib is administered orally, usually in a dose of 50 mg daily in cycles of 4 weeks on, then 2 weeks off. However, a lower dose of 37.5 mg daily is also being used, since patients tend to tolerate the lower dosage better, and tumor growth has sometimes been documented during off weeks with the cycled regimen. The dosage may be adjusted in 12.5-mg increments according to response and toxicity. Sunitinib is metabolized in the liver by CYP3A4, so dose adjustments should be made for patients taking the wide variety of drugs that are inhibitors or inducers of CYP3A4. Sunitinib can cause serious adverse reactions, including heart failure, cardiac arrhythmias, marrow suppression, pancreatitis, hypo- or hyperthyroidism, nephrotic syndrome, and rhabdomyolysis with acute renal failure. Patients treated with sunitinib also commonly experience nausea, vomiting, diarrhea, hypertension, skin discoloration, mucositis, asthenia, dyspnea, myalgias, and arthralgias.

Adjuvant statin therapy: Certain lipophilic statins (lovastatin, fluvastatin, and simvastatin) inhibit the growth and migration of mouse PHEO cells and the more aggressive mouse PHEO tissue-derived cells *in vitro*. The effect appears to be mediated through inhibition of mitogen-activated kinase (MAPK) pathway. The MAPK pathway appears to play a role in certain human PHEO/PGL tumors with mutations in *K-RAS, RET, NF-1,* and *SDHB*. Although oral administration of these statins would not achieve adequate tumor levels to cause tumor cell apoptosis, the use of these statins is a promising therapeutic option as an adjuvant to chemotherapy. However, there have been no clinical trials to indicate whether this effect is clinically significant.

Potential therapeutic targets: ATP synthase (ATP5B) is an enzyme complex that is intracellular in most normal cells but is found on the surface of SDHB-associated PGLs and may promote tumor cell survival. Inhibition of tumoral ATP synthase with anti-ATP synthase antibodies and drugs that inhibit ATP synthase (eg, resveratrol) are possible future therapies against metastatic PHEO/PGL.

Triptolide, an inhibitor of nuclear factor-kappaB (NF-κB) increases the expression of norepinephrine transporter (NET) in PHEO cells and has increased the uptake of MIBG into mouse PHEO cells *in vitro*. Triptolide has also reduced the tumor burden in a metastatic PHEO animal model. Triptolide has serious clinical adverse effects; a better-tolerated prodrug (minnelide) is in early clinical trials for patients with metastatic GI tumors.

D. Bisphosphonates or denosumab   No controlled clinical trials have assessed the efficacy of bisphosphonates or

denosumab against osteolytic bone metastases from PGLs and PHEOs. However, bisphosphonates and denosumab have demonstrated effectiveness in other osteolytic solid tumors to reduce skeletal-related adverse events. Zoledronic acid is usually administered in doses of 4 to 5 mg every 1 to 2 months as an intravenous infusion to patients with osteolytic bone metastases. Patients unable to tolerate zoledronic acid may tolerate intravenous pamidronate. Denosumab is administered subcutaneously at initial doses of 60 mg every 1 to 2 months for patients with osteolytic bone metastases. Long-term therapy with either agent is associated with an increased risk for atypical subtrochanteric (chalk stick) femoral fractures, aseptic necrosis of the jaw, and metatarsal stress fractures.

E. Radiation therapy   When administered to patients with symptomatic spinal or cranial PHEO/PGL metastases, radiation therapy can reduce pain and produce neurologic improvement. Although conventional radiation therapy is usually administered, gamma knife can be given to smaller symptomatic bone metastases and to jugulotympanic PGLs. Conventional radiation therapy to large primary tumors or intra-abdominal metastases is not advisable, because it is usually ineffective and causes morbidity such as radiation enteritis and a proclivity to later surgical complications such as wound dehiscence, infections, and fistulas. However, small recurrent tumors can be treated with CyberKnife stereotactic radiosurgery. Surgical debulking of large abdominal or thoracic tumors (or other therapies) is usually preferable to radiation therapy. Radiation therapy to tumors reduces their uptake of ^{131}I-MIBG.

F. Arterial embolization and radiofrequency ablation   Before arterial embolization or radiofrequency ablation, patients with secretory PHEO/PGL must be fully prepared such that their blood pressure is near normal as described earlier. Pretreatment includes α-blockade and/or other measures such as β-blockade, CCBs, or metyrosine. Patients are monitored with arterial blood pressure transducers and given a central line before endotracheal general anesthesia. Anesthesia standby is necessary in case severe hypertension occurs, and it becomes necessary to administer intravenous antihypertensive drugs.

Arterial embolization has been used on rare occasions to reduce blood flow to PGLs and PHEOs, either in preparation for surgery or for inoperable cases. These tumors are usually very vascular and preoperative embolization may reduce intraoperative hemorrhage. Also, embolizing the tumor's blood supply may slow its growth. However, there have been no controlled clinical trials as to its effectiveness. A major potential risk of embolization is that of PHEO crisis. However, embolization has been used successfully on secretory tumors where the patient has been fully prepared with α-blockade and/or other measures. Localization arteriograms should use nonionic contrast.

Radiofrequency (RF) thermal ablation has been used successfully for patients with metastatic PHEO/PGL, particularly for liver and bone metastases. Most reported RF ablations of bone metastases have targeted rib or ischial/pelvic lesions. Before RF

ablations, the RF electrode(s) is guided into the tumor with ultrasound or CT guidance. Single electrodes may be used for small lesions, while metastases over 2.5 cm diameter usually require triple parallel cluster needle electrodes.

G. ^{131}I-metaiodobenzylguanidine (^{131}I-MIBG) (Figure 11–14) ^{131}I-MIBG is a treatment option for patients with metastatic or unresectable PHEO or PGL whose tumors are avid for the isotope. Benzylguanidine is a false neurotransmitter that resembles norepinephrine and is preferentially absorbed by PHEO/PGL tumors that express surface norepinephrine transporter 1 (NET 1) activity; it is tagged with ^{131}I to produce ^{131}I-MIBG. Only about 60% of metastatic or recurrent unresectable PHEO/PGLs have sufficient avidity for MIBG for ^{131}I-MIBG therapy to be potentially efficacious. Nonsecretory PHEO/PGLs frequently have sufficient MIBG to allow treatment. Pretreatment with nifedipine increased MIBG retention in 4/8 patients in one report. Some patients with good uptake of MIBG on diagnostic scanning have disappointingly poor uptake of large therapeutic doses of ^{131}I-MIBG, possibly caused by competitive inhibition by large amounts of nonradioactive ("carrier added") ^{127}I-MIBG that are present in most current formulations of ^{131}I-MIBG. A "no-carrier-added" formulation of ^{131}I-MIBG has been developed and is in clinical trials. Minnelide, a prodrug for an inhibitor of NF-κB has the potential for possibly increasing MIBG uptake into PHEO/PGL tumors (discussed later). Responses to ^{131}I-MIBG therapy occur over several months and continued decreases in tumor markers and tumor size may be observed for up to a year following therapy.

Protocol: Treatment protocols vary among institutions. Most currently employ repeated low-activity therapies of 100 to 250 mCi (3.7-9.25 GBq) or intermediate activity therapies of 250 to 500 mCi (9.25-18.5 GBq). High-activity ^{131}I-MIBG therapies over 500 mCi (18.5 GBq) have been employed for metastatic PHEO/PGL, but pose increased risks. ^{131}I-MIBG is administered intravenously over 30 to 90 minutes via a peripheral or central venous catheter. Following therapy with ^{131}I-MIBG, once background radiation has dissipated, a posttreatment whole-body scan is obtained.

Before and after ^{131}I-MIBG therapy, patients with an intact thyroid are premedicated with oral potassium iodide to reduce the risk of thyroid damage that could be caused by free ^{131}I generated through metabolism of ^{131}I-MIBG. The best-tolerated oral potassium iodide preparation is ThyroShield. Patients are all pretreated with non-phenothiazine antiemetics such as ondansetron; they remain hospitalized in lead-shielded rooms with radiation safety precautions until the emitted gamma radiation declines to acceptable levels, which usually requires about 4 to 9 days, depending upon the amount of activity infused and its uptake and retention by PHEO/PGL tumors. Repeated treatments may be required.

Responses: Complete remission (CR) or partial remission (PR), defined by RECIST (Response Evaluation Criteria in Solid Tumors), occur in about 22% of patients treated with ^{131}I-MIBG, while an additional 35% have minor responses (improvement in symptoms, tumor markers, and tumor size not meeting RECIST
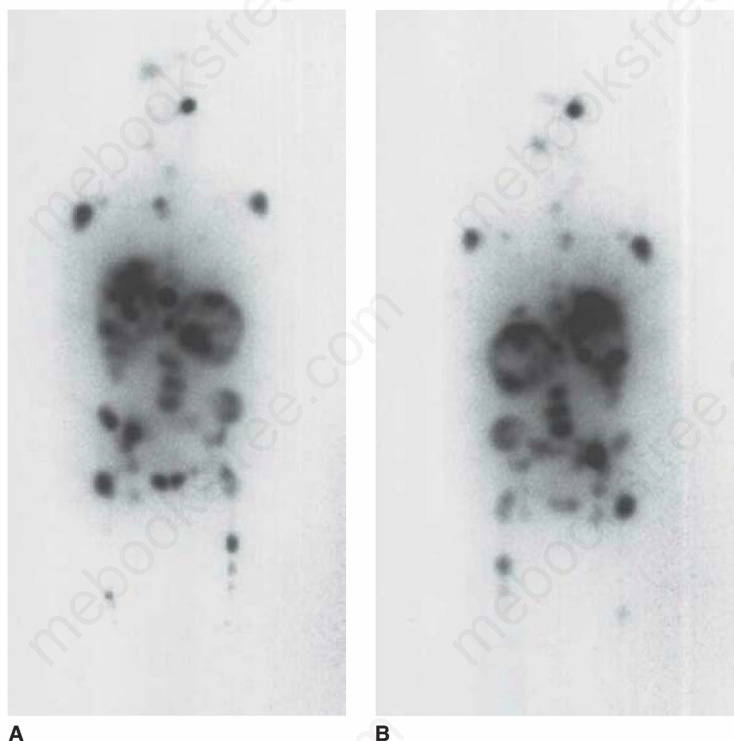
FIGURE 11–14 Posttherapy scan 1 week after administration of 500 mCi ^{131}I-MIBG. Patient is a 22-year-old woman with a large unresectable PHEO in the left adrenal that has central hemorrhagic necrosis. There are widespread metastases to liver and bones. A metastasis to the left orbit has caused optic nerve compression and a visual field defect. A large metastasis to the left pelvis has caused pain and required prior external beam radiation therapy. **A.** Anterior image; **B.** Posterior image.

criteria for a PR); another 8% have stable disease (SD), at least temporarily. About one-third of patients fail to respond at all to ^{131}I-MIBG and experience progressive disease (PD) within a year after therapy. ^{131}I-MIBG therapy may improve 5-year survival, but controlled clinical trials are lacking.

Adverse reactions: Intermediate or high-activity ^{131}I-MIBG therapy predictably causes immediate lymphocytopenia, followed by some degree of platelet and neutrophil suppression that commences about 2.5 to 4 weeks after therapy. Therefore, patients must be observed and treated for any thrush or inguinal candidiasis and complete blood counts with platelet counts must be monitored after therapy. Platelet transfusions, red blood cell transfusions, or filgrastim may be required. High-activity therapy with ^{131}I-MIBG greater than 500 mCi (18.5 GBq) risks prolonged bone marrow suppression, requiring pretreatment peripheral blood stem cell harvesting with cryopreservation. High-activity ^{131}I-MIBG also increases radiation exposure to hospital personnel and is especially problematic if the patient develops complications and requires care in an intensive care unit without lead shielding.

Other adverse reactions to ^{131}I-MIBG therapy include nausea, occasional sialadenitis, transient hair loss, hypogonadism, and infertility. Also, ARDS and multisystem crisis have occurred in rare patients with proteinuria (a marker for this complication) after ^{131}I-MIBG therapy with activities greater than or equal to 500 mCi (18.5 GBq). Patients with PHEO/PGL who have any degree of proteinuria are at increased risk for ARDS and multisystem crisis, either spontaneously or after ^{131}I-MIBG or other

procedures. Patients with metastatic PHEO/PGL and proteinuria may be treated with repeated low-activity ^{131}I-MIBG therapies. Patients treated with ^{131}I-MIBG have an increased lifetime risk of second malignancies, particularly myelodysplastic syndrome (MDS) and leukemia. MDS and leukemia have occurred about 5 to 7 years after ^{131}I-MIBG, particularly in patients who have received cumulative activities greater than or equal to 1000 mCi (37 GBq).

Prognosis

The mortality rate for patients undergoing PHEO/PGL resection has dropped to under 3% due to improved medical preparation and surgical technique. Laparoscopic resections have reduced perioperative morbidity and shortened the length of hospitalization. However, even after complete resection of the PHEO/PGL, hypertension persists or recurs in 25% of patients. Recurrent hypertension is an indication for reevaluation for PHEO.

Following surgical resection of a PHEO/PGL without detectable metastases, patients have a 5-year survival rate of 96%. Risk factors for death from PHEO/PGL include tumor size over 5 cm, metastatic disease, and local tumor invasion. However, the long-term mortality rate is much higher than expected. In a Swedish long-term outcome study of 121 patients with PHEO/PGL, there was no perioperative mortality, but 50% of patients remained hypertensive postoperatively. Of the 121 patients, 42 died during an observation period averaging 15 years, versus an expected

24 deaths in an age-matched control population. Thus, their relative risk of mortality was increased 78% (RR 1.78). Of the 42 patients who died, 20 deaths were due to cardiovascular disease, 6 from associated neuroectodermal tumors, 5 from other malignancies, 7 from unrelated causes, and 4 from malignant PHEO/PGL.

Metastases are variably aggressive, between patients and within a given patient. Some metastases are quite indolent and may present clinically one or two decades after resection of the primary tumor. However, other metastases are exceptionally aggressive. Asymptomatic patients with only a few bone metastases tend to have the best prognosis, while those with a heavy burden of liver and lung metastases tend to have the most malignant disease.

Patients with metastatic PHEO/PGL who have appropriate surveillance and treatment now have average overall survival rates as follows: 5-year survival, 77%; 10-year survival, 62%; 20-year survival, 39%. *SDHB* germline mutations are found in about 81% of children with metastatic PHEOs/PGLs, compared to about 49% of adults. Genetics affects prognosis. Adults *SDHB* germline mutations tend to fare worse than patients with metastases from sporadic PHEO/PGL. However, children with metastatic PHEO/PGL associated with *SDHB* germline mutations tend to fare better than their adult counterparts.

PHEO and PGL: Postoperative Long-Term Surveillance

A preoperative ^{123}I-MIBG scan or other radionuclide scan (discussed earlier) is recommended for all patients with PHEO/PGL. Postoperative scanning is recommended for patients when there is any doubt about complete resection of an adrenal PHEO and for any patients with PGL or multiple tumors. The postoperative scan is usually obtained several months after surgery.

All patients with PHEOs/PGLs require lifetime postoperative surveillance. They require aggressive treatment of all cardiovascular risk factors, due to their increased long-term risk of death from cardiovascular causes. Additionally, patients should be tested for familial genetic syndromes and appropriately screened for associated malignancies (see Genetic Conditions Associated with Pheochromocytomas and Paragangliomas, see Tables 11–8 through 11–12). Persistent symptoms or hypertension can signify recurrence at the surgical site, seeding of the peritoneum, a contralateral PHEO, a PGL, or possibly metastatic disease. About 10% of adrenal PHEOs have metastasized at the time of diagnosis or soon postoperatively. However, occult metastatic disease is detected up to 20 years later in another 5%. Up to 35% of sympathetic PGLs have metastases at the time of diagnosis, particularly in patients harboring *SDHB* or fumarate hydratase (*FH*) germline mutations. Other patients develop multiple recurrent intra-abdominal tumors (pheochromocytomatosis) probably caused by tumor seeding that may occur spontaneously from the original tumor or during surgery.

Patients with secretory tumors are usually followed with plasma-fractionated free metanephrine determinations. Plasma-fractionated catecholamines and dopamine may also be obtained if they were predominantly secreted by the primary tumor. Serum CgA is a useful tumor marker for patients with PHEOs/PGLs

whose primary tumor secreted CgA and whose renal function is normal; elevated and rising levels of CgA usually indicate tumor recurrence or metastases (see CgA, discussed earlier). The type of biochemical follow-up is tailored to the individual patient. The first determination of postoperative plasma-fractionated free metanephrines is obtained at least 2 weeks after surgery, because catecholamine excretion often remains high for up to 10 days after successful surgery. Testing is obtained quarterly during the first year following surgery, then semiannually for at least 5 years. After 5 years, lifetime routine yearly physical examinations and biochemical screening are recommended, with immediate evaluation for recurrent PHEO/PGL if suspicious symptoms recur. For hypertensive patients, weekly home blood pressure monitoring is recommended for the first year postoperatively and monthly thereafter. A rising blood pressure or recurrence of symptoms should trigger a full workup for recurrent or metastatic PHEO/PGL.

REFERENCES

Adrenal Medulla and Adrenergic Physiology

Biaggioni I, Robertson D. Adrenoceptor agonists and sympathomimetic drugs. In: Katzung BG, ed. *Basic and Clinical Pharmacology*. 12th ed. New York: McGraw Hill; 2012.

Tank AW, Lee Wong D. Peripheral and central effects of circulating catecholamines. *Compr Physiol*. 2015;5:1. [PMID: 25589262]

Autonomic Insufficiency

Bornstein SR, Breidert M, Ehrhart-Bornstein M, Kloos B, Scherbaum WA. Plasma catecholamines in patients with Addison's disease. *J Clin Endocrinol (Oxf)*. 1995;42:215. [PMID: 7704967]

Kaufmann H, Norcliffe-Kauffman L, Palma JA. Droxidopa in neurogenic orthostatic hypotension. *Expert Rev Cardiovasc Ther*. 2015;13:875. [PMID: 26092297]

Loaverbruck A, Sandroni P. Neurogenic orthostatic hypotension: roles of norepinephrine deficiency in its causes, its treatment, and future research directions. *Curr Med Res Opin*. 2015;31:2095. [PMID: 26373628]

Pheochromocytoma and Paranglioma Reviews

Kiernan CM, Solórzano CC. Pheochromocytoma and paraganglioma: diagnosis, genetics, and treatment. *Surg Oncol Clin N Am*. 2016;25:119. [PMID: 26610778]

Lai EW, Perera SM, Havekes B, et al. Gender-related differences in the clinical presentation of malignant and benign pheochromocytoma. *Endocrine*. 2008;34:96. [PMID: 18982461]

Lenders JW, Duh QY, Eisenhofer G, et al. Pheochromocytoma and paraganglioma: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:1915. [PMID: 24893135]

Pacak K, Wimalawansa SJ. Pheochromocytoma and paraganglioma. *Endocr Pract*. 2015;21:406. [PMID: 25716634]

Sutton MG, Sheps SG, Lie JT. Prevalence of clinically unsuspected pheochromocytoma. Review of a 50-year autopsy series. *Mayo Clin Proc*. 1981;56:354. [PMID: 6453259]

Pheochromocytoma and Paraganglioma Pathophysiology

Abdel-Aziz E, Prete F, Conway G, et al. Phaeochromcytomas and paragangliomas: a difference in disease behavior and clinical outcomes. *J Surg Oncol*. 2015;112:486. [PMID: 26337393]

Bugalho MJ, Silva AL, Domingues R. Coexistence of paraganglioma/pheochromocytoma and papillary thyroid carcinoma: a four-case series analysis. *Fam Cancer*. 2015;14:603. [PMID: 26071763]

Opotowsky AR, Moko LE, Ginns J, et al. Pheochromocytoma and paraganglioma in cyanotic congenital heart disease. *J Clin Endocrinol Metab*. 2015;100:1325. [PMID: 25581599]

Thosani S, Ayala-Ramirez M, Román-González A, et al. Constipation: an overlooked, unmanaged symptom of patients with pheochromocytoma and sympathetic paraganglioma. *Eur J Endocrinol*. 2015;173(3):377. [PMID: 26060051]

Tischler AS, del Krijger RR. 15 years of paraganglioma: pathology of pheochromocytoma and paraganglioma. *Endocr Relat Cancer*. 2015;22:T123. [PMID: 26136457]

Genetic Aspects of Pheochromocytoma and Paraganglioma

Astrom K, Cohen JE, Willett-Brozick JE, et al. Altitude is a phenotypic modifier in hereditary paraganglioma type 1: evidence for an oxygen-sensing defect. *Hum Genet*. 2003;113:228. [PMID: 12811540]

Benn DE, Robinson BG, Clifton-Bligh RJ. 15 years of paraganglioma: clinical manifestations of paraganglioma syndromes types 1-5. *Endocr Relat Cancer*. 2015;22:T91. [PMID: 26273102]

Boikos SA, Xekouki P, Fumagalli E, et al. Carney triad can be (rarely) associated with germline succinate dehydrogenase defects. *Eur J Hum Genet*. 2016;24:569. [PMID: 26173966]

Burnichon N, Buffet A, Gimenez-Roqueplo AP. Pheochromocytoma and paraganglioma: molecular testing and personalized medicine. *Curr Opin Oncol*. 2016;28:5. [PMID: 26599293]

Carney JA. Carney triad. *Front Horm Res*. 2013;41:92. [PMID: 23652673]

Cascón A, Comino-Méndez I, Currás-Freixes M, et al. Whole-exome sequencing identifies MDH2 as a new familial paraganglioma gene. *J Natl Cancer Inst*. 2015;107:1-5. [PMID: 25766404]

Cassol C, Mete O. Endocrine manifestations of von Hippel-Lindau disease. *Arch Pathol Lab Med*. 2015;139:263. [PMID: 25611110]

Comino-Méndez I, de Cubas AA, Bernal C, et al. Tumoral EPAS1 (HIF2A) mutations explain sporadic pheochromocytoma and paraganglioma in the absence of erythrocytosis. *Hum Mol Genet*. 2013;22:2169. [PMID: 23418310]

Currás-Freixes M, Inglade-Pérez L, Mancikova V, et al. Recommendations for somatic and germline genetic testing of single pheochromocytoma and paraganglioma based on findings from a series of 329 patients. *J Med Genet*. 2015;52:647. [PMID: 26269449]

Fishbein L. Pheochromocytoma and paraganglioma: genetics, diagnosis, and treatment. *Hematol Oncol Clin North Am*. 2016;30:135. [PMID: 26614373]

Hernandez KG, Ezzat S, Morel CF, et al. Familial pheochromocytoma and renal cell carcinoma syndrome: TMEM127 as a novel candidate gene for the association. *Virchows Arch*. 2015;466:727. [PMID: 25800244]

Ladroue C, Carcenac R, Leporrier M, et al. PHD2 mutation and congenital erythrocytosis with paraganglioma. *N Engl J Med*. 2008;359:25. [PMID: 19092153]

Lang BH, Yu HW, Lo Cy, et al. Bilateral pheochromocytomas in MEN2A syndrome: a two-institution experience. *World J Surg*. 2015;39:2484. [PMID: 26071011]

Lefebvre M, Foulkes WD. Pheochromocytoma and paraganglioma syndromes: genetics and management update. *Current Oncology*. 2014;21:e8. [PMID: 24523625]

Lorenzo FR, Yang C, Ng Tang Fui M, et al. A novel EPAS1/HIF2A germline mutation in a congenital polycythemia with paraganglioma. *J Mol Med (Berl)*. 2013;91:507. [PMID: 23090011]

Martins R, Bugalho MJ. Paragangliomas/pheochromocytomas: clinically oriented genetic testing. *Int J Endocrinol*. 2014;2014:794187. [PMID: 24899893]

Neumann HP, Vortmeyer A, Schmidt D, et al. Evidence of MEN-2 in the original description of classic pheochromocytoma. *N Engl J Med*. 2007;357:1311. [PMID: 17898100]

Pęczkowska M, Kowalska A, Sygut J, et al. Testing new susceptibility genes in the cohort of apparently sporadic phaeochromocytoma/paraganglioma patients with clinical characteristics of hereditary syndromes. *Clin Endocrinol (Oxf)*. 2013;79:817. [PMID: 23551045]

Toledo RA, Dahia PL. Next-generation sequencing for the diagnosis of hereditary pheochromocytoma and paraganglioma syndromes. *Curr Opin Endocrinol Diabetes Obes*. 2015;22:169. [PMID: 25871962]

Xekouki P, Szarek E, Bullova P, et al. Pituitary adenoma and paraganglioma/pheochromocytoma (3Ps) and succinate dehydrogenase defects in humans and mice. *J Clin Endocrinol Metab*. 2015;100:E710. [PMID: 25695889]

Yang C, Zhuang Z, Fliedner SM, et al. Germ-line PHD1 and PHD2 mutations detected in patients with pheochromocytoma/paraganglioma polycythemia. *J Mol Med (Berl)*. 2015;93:93. [PMID: 25263965]

Zhuang Z, Yang C, Lorenzo F, et al. Somatic HIF2A gain-of-function mutations in paraganglioma with polycythemia. *New Engl J Med*. 2012;367:922. [PMID: 22931260]

Diagnostic Tests and Differential Diagnosis for Pheochromocytoma

Eisenhofer G, Peitzsch M. Laboratory evaluation of pheochromocytoma and paraganglioma. *Clin Chem*. 2014;60:1486. [PMID: 25332315]

Sethi PS, Hiser W, Gaffar H, et al. Fatal pheochromocytoma crisis precipitated by dobutamine stress echocardiography. *Eur J Echocardiogr*. 2008;9:819. [PMID: 18579490]

Vanderveen KA, Thompson SM, Callstrom MR, et al. Biopsy of pheochromocytoma and paraganglioma: potential for disaster. *Surgery*. 2009;146:1158. [PMID: 19958944]

Yu R, Weu M. False positive test results for pheochromocytoma from 2000 to 2008. *Exp Clin Endocrinol Diabetes*. 2010;118:577. [PMID: 19998239]

Localizing Scans for Pheochromocytoma and Paraganglioma

Allen BC, Francis IR. Adrenal imaging and intervention. *Radiol Clin North Am*. 2015;53:1021. [PMID: 26321451]

Janssen I, Blanchet EM, Adams K, et al. Superiority of [68Ga]-DOTATATE PET/CT to other functional imaging modalities in the localization of SDHB-associated metastatic pheochromocytoma and paraganglioma. *Clin Cancer Res*. 2015;21:3888. [PMID: 25873086]

Koopmans KP, Jager PL, Kema IP, Kerstens MN, Albers F, Dullaart RP. ^{111}In-octreotide is superior to ^{123}I-metaiodobenzylguanidine for scintigraphic detection of head and neck paragangliomas. *J Nucl Med*. 2008;48:1232. [PMID: 18632829]

Oh JR, Kulkani H, Carreras C, et al. Ga-68 somatostatin receptor PET/CT in von Hippel-Lindau disease. *Nucl Med Mol Imaging*. 2012;46:129. [PMID: 24900047]

Tan TH, Hussein Z, Saad FF, Shualb IL. Diagnostic performance of (68)Ga-DOTATATE PET/CT, (18)F-FDG PET/CT and (131)I-MIBG scintigraphy in mapping metastatic pheochromocytoma and paraganglioma. *Nucl Med Mol Imaging*. 2015;49:143. [PMID: 26085860]

Incidental Adrenal Tumors

Heavner MG, Krane LS, Winters SM, Mirzazadeh M. Pheochromocytoma diagnosed pathologically with previous negative serum markers. *J Surg Oncol*. 2015;112:492. [PMID: 26384104]

Ioachimescu AG, Remer EM, Hamrahian AH. Adrenal incidentalomas: a disease of modern technology offering opportunities for improved patient care. *Endocrinol Metab Clin North Am*. 2015;44:335. [PMID: 26038204]

Jun JH, Ahn HJ, Lee SM, et al. Is preoperative biochemical testing for pheochromocytoma necessary for all adrenal adenomas? *Medicine (Baltimore)*. 2015;94:e1948. [PMID: 26559265]

Loh HH, Yee A, Loh HS, et al. The natural progression and outcomes of adrenal incidentaloma: a systematic review and meta-analysis. *Minerva Endocrinol*. 2017;42(1):77. [PMID: 26698544]

Mazzaglia PJ, Monchik JM. Limited value of adrenal biopsy in the evaluation of adrenal neoplasm: a decade of experience. *Arch Surg*. 2009;144:465. [PMID: 19451490]

Medical Therapy for Pheochromocytoma

Habbe N, Ruger F, Bojunga J, Beckstein WO, Holzer K. Urapidil in the preoperative treatment of pheochromocytoma. A safe and cost-effective method. *World J Surg*. 2013;37:1141. [PMID: 22381676]

Kuok CH, Yen CR, Huang CS, Ko YP, Tsai PS. Cardiovascular collapse after labetalol for hypertensive crisis in an undiagnosed pheochromocytoma during cesarean section. *Acta Anaesthesiol Taiwan*. 2011;47:69. [PMID: 21729814]

Thosani S, Ayala-Ramirez M, Román-González A, et al. Constipation: an over-looked, unmanaged symptom of patients with pheochromocytoma and sympathetic paraganglioma. *Eur J Endocrinol.* 2015;173:377. [PMID: 26060051]

Vendrell M, Martin N, Tejedor A, et al. Magnesium sulphate and (123)I-MIBG in pheochromocytoma: two useful techniques for a complicated disease. *Rev Esp Anestesiol Reanim.* 2015;63:48. [PMID: 26025287]

Westaby S, Shahir A, Sadler G, Flynn F, Ormerod O. Mechanical bridge to recovery in pheochromocytoma myocarditis. *Nat Rev Cardiol.* 2009;6:482. [PMID: 19554007]

Pheochromocytoma in Pregnancy

Kiroplastis K, Kambaroudis A, Andronikou A, et al. Dealing with pheochromocytoma during the first trimester of pregnancy. *Case Rep Obstet Gynecol.* 2015;2015:439127. [PMID: 25838955]

Petrie J, Lockie C, Paolineli A, et al. Undiagnosed phaeochromocytoma masquerading as eclampsia. *BMJ Case Rep.* 2012;2012. [PMID: 22665869]

Prete A, Paragliola RM, Salvatori R, Corsello SM. Management of catecholamine-secreting tumors in pregnancy: a review. *Endocr Pract.* 2016;22:357. [PMID: 26536138]

Wing LA, Conaglen JV, Meyer-Rochow GY, Elston MS. Paraganglioma in pregnancy: a case series and review of the literature. *J Clin Endocrinol Metab.* 2015;100:3202. [PMID: 26083822]

Pheochromocytoma-Induced Life-Threatening Complications: Cardiomyopathy, Acute Respiratory Distress Syndrome (ARDS), and Multisystem Crisis

Barrett C, van Uum SH, Lenders JW. Risk of catecholaminergic crisis following glucocorticoid administration in patients with an adrenal mass: a literature review. *Clin Endocrinol (Oxf).* 2015;83:622. [PMID: 25940577]

Batissie-Lignier M, Pereira B, Motreff P, et al. Acute and chronic pheochromocytoma–induced cardiomyopathy: different prognoses?: a sympathethc analytic review. *Medicine (Baltimore).* 2015;95:e2198. [PMID: 26683930]

Hosseinnezhad A, Black RM, Aeddula NR, Adhidari D, Trivedi N. Glucagon-induced pheochromocytoma crisis. *Endocr Pract.* 2011;17:e51. [PMID: 21324811]

Kakoki K, Miyata Y, Shida Y, et al. Pheochromocytoma multisystem crisis treated with emergency surgery: a case report and literature review. *BMC Res Notes.* 2015;8:758. [PMID: 26645353]

Otusana O, Goraya H, Iyer P, Landi K, Tibb A, Msaouel P. A vicious cycle of acute catecholamine cardiomyopathy and circulatory collapse secondary to pheochromocytoma. *Oxf Med Case Reports.* 2015;2015:343. [PMID: 26512333]

Porzig A, Matthay KK, Dubois S, et al. Proteinuria in metastatic pheochromocytoma is associated with an increased risk of acute respiratory distress syndrome, spontaneously or after therapy with ^{131}I-meta-iodobenzylguanidine. *Horm Metab Res.* 2012;44:539. [PMID: 22588707]

Rassler B. Role of α- and β-adrenergic mechanism in the pathogenesis of pulmonary injuries characterized by edema, inflammation, and fibrosis. *Cardiovasc Hematol Disord Drug Targets.* 2013;13:197. [PMID: 24479719]

Salahuddin A, Rohr-Kirchgraber T, Shekar R, West B, Loewenstein J. Interleukin-6 in the fever and multiorgan crisis of pheochromocytoma. *Scand J Infect Dis.* 1997;29:640. [PMID: 9571753]

Takahashi N, Shimada T, Tanabe K, Yoshitomi H, et al. Steroid-induced crisis and rhabdomyolysis in a patient with pheochromocytoma: case report and review. *Int J Cardiol.* 2011;146:e41. [PMID: 19187985]

Yoshikawa T. Takotsubo cardiomyopathy, a new concept of cardiomyopathy: clinical features and pathophysiology. *Int J Cardiol.* 2015;182:297. [PMID: 25585367]

Surgery for Pheochromocytoma and Paraganglioma

Brunaud L, Nguyen-Thi PL, Mirallie E, et al. Predictive factors for postoperative morbidity after laparoscopic adrenalectomy: a multicenter retrospective analysis in 225 patients. *Surg Endosc.* 2016;30:1051. [PMID: 26092019]

Bruynzeel H, Feelders RA, Groenland TH, et al. Risk factors for hemodynamic instability during surgery for pheochromocytoma. *J Clin Endocrinol Metab.* 2010;95:678. [PMID: 19965926]

Castinetti F, Taieb D, Henry JF, et al. Management of endocrine disease: outcome of adrenal sparing surgery in heritable pheochromocytoma. *Eur J Endocrinol.* 2016;174:R9. [PMID: 26297495]

Colleselli D, Janetschek G. Current trends in partial adrenalectomy. *Curr Opin Urol.* 2015;25:89. [PMID: 25581540]

Gaujoux S, Bonnet S, Lentschener C, et al. Preoperative risk factors of hemodynamic instability during laparoscopic adrenalectomy for pheochromocytoma. *Surg Endosc.* 2016;30:2984. [PMID: 26684206]

Hotu C, Harman R, Cutfield R, et al. Laparoscopic adrenalectomy for phaeochromocytoma: a case series. *N Z Med J.* 2015;128:35. [PMID: 26645753]

Kanazawa S, Uehara N, Seo H, Miyawaki I, Yamazaki K. Low dose vasopressin is effective for catecholamine-resistant hypotension after resection of pheochromocytoma. *Masui.* 2013;62:1218. [PMID: 24228460]

Ramakrishna H. Pheochromocytoma resection: current concepts in anesthetic management. *J Anaesthesiol Clin Pharmacol.* 2015;31:317. [PMID: 26330708]

Radiation, Embolization and Radiofrequency Ablation of Paragangliomas, Pheochromocytomas, and Metastases

Huy PT, Kania R, Duet M, Dessard-Diana B, Mazeron JJ, Benhamed R. Evolving concepts in the management of jugular paraganglioma: a comparison of radiotherapy and surgery in 88 cases. *Skull Base.* 2009;19:83. [PMID: 19568345]

Liscak R, Urgosik D, Chytka T, et al. Leksell gamma knife radiosurgery of the jugulotympanic glomus tumomr: long–term results. *J Neurosurg.* 2014;121:S198. [PMID: 25434953]

Venkatesan AM, Locklin J, Lai EW, et al. Radiofrequency ablation of metastatic pheochromocytoma. *J Vasc Interv Radiol.* 2009;20:1483. [PMID: 19875067]

Vogel J, Atanacio AS, Prodanov T, Turkbey BI, Adams K, Martucci V, et al. External beam radiation therapy in treatment of malignant pheochromocytoma and paraganglioma. *Front Oncol.* 2014;4:166. [PMID: 25019060]

Metastatic Pheochromocytoma, Paraganglioma, and Pheochromocytomatosis

Blake GM, Lewington VJ, Fleming JS, et al. Modification by nifedipine of ^{131}I-meta-iodobenzylguanidine kinetics in malignant pheochromocytoma. *Eur J Nucl Med.* 1988;14:345. [PMID: 3181183]

Fliedner SM, Engel T, Lendval NK, et al. Anti-cancer potential of MAPK pathway inhibition in paragangliomas—effect of different statins on mouse pheochromocytoma cells. *PLOS One.* 2014;9:e97712. [PMID: 24846270]

Fliedner SM, Yang C, Thompson E, et al. Potential therapeutic target for malignant paragangliomas: ATP synthase on the surface of paraganglioma cells. *Am J Cancer Res.* 2015;15:1558. [PMID: 26101719]

Gonias S, Goldsby R, Matthay KK, et al. Phase II study of high-dose [^{131}I] metaiodobenzylguanidine therapy for patients with metastatic pheochromocytoma and paraganglioma. *J Clin Oncol.* 2009;27:6162. [PMID: 19636009]

Hadoux J, Favier J, Scoazec JY, et al. SDHB mutations are associated with response to temozolomide in patients with metastatic pheochromocytoma or paraganglioma. *Int J Cancer.* 2014;135:2711. [PMID: 24752622]

Hescot S, Leboulleux S, Amar L, Vezzosi D, Borget I, Bournaud-Salinas C, et al. One-year progression-free survival of therapy-naïve patients with malignant pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab.* 2013;9:4006. [PMID: 234775]

Huang H, Abraham J, Hung E, et al. Treatment of malignant pheochromocytoma/paraganglioma with cyclophosphamide, vincristine, and dacarbazine: recommendation from a 22-year follow-up of 18 patients. *Cancer.* 2008;113:2020. [PMID: 18780317]

Joshua AM. Rationale and evidence for sunitinib in the treatment of malignant paraganglioma/pheochromocytoma. *J Clin Endocrinol Metab.* 2009;94:5. [PMID: 19001511]

Niemeijer ND, Alblas G, van Hulsteijn LT, Dekkers OM, Corssmit EP. Chemotherapy with cyclophosphamide, vincristine and dacarbazine for malignant paraganglioma and pheochromocytoma: systematic review and meta-analysis. *Clin Endocrinol (Oxf).* 2014;81:642. [PMID: 25041164]

Pacak K, Sirova M, Giubellino A, et al. NF-κB inhibition significantly upregulates the norepinephrine transporter system, causes apoptosis in pheochromocytoma cell lines and prevents metastasis in an animal model. *Int J Cancer.* 2012;131:2445. [PMID: 22407736]

Rutherford MA, Rankin AJ, Yates TM, et al. Management of metastatic phaeo-chromocytoma and paraganglioma: use of iodine-131-metaiodobenzylguani-dine therapy in a tertiary referral centre. *QJM.* 2015;108:361. [PMID: 25267727]

Turkova H, Prodanov T, Maly M, et al. Characteristics and outcomes of metastatic SDHB and sporadic pheochromocytoma/paraganglioma: an NIH study. *Endocr Pract.* 2016;22:302. [PMID: 26523625]

Yoshinaga K, Oriuchi N, Wakabayashi H, et al. Effects and safety of ^{131}I-metaiodo-benzylguanidine (MIBG) radiotherapy in malignant neuroendocrine tumors: results from a multicenter observational registry. *Endocr J.* 2014;61:1171. [PMID: 25214026]

Testes

Bradley D. Anawalt, MD and Glenn D. Braunstein, MD

ACTH	Adrenocorticotropic hormone	**IVF**	*In vitro* fertilization
cAMP	Cyclic adenosine monophosphate	**LH**	Luteinizing hormone
DHEA	Dehydroepiandrosterone	**mRNA**	Messenger ribonucleic acid
DHT	Dihydrotestosterone	**PDE5**	Type 5 phosphodiesterase
FSH	Follicle-stimulating hormone	**PRL**	Prolactin
GnRH	Gonadotropin-releasing hormone	**SHBG**	Sex hormone–binding globulin
hCG	Human chorionic gonadotropin	**TGFβ**	Transforming growth factor β
ICSI	Intracytoplasmic sperm injection		

The testes contain two major components that are structurally separate and serve different functions. The **Leydig cells**, or **interstitial cells**, produce sex steroid hormones and comprise one of the two major endocrine cells of the testes. The primary secretory product of Leydig cells, testosterone, is responsible either directly or indirectly for embryonic differentiation along male lines of the external and internal genitalia, male secondary sexual development at puberty, and maintenance of libido and potency in the adult male. The Leydig cells make 4 to 6 mg of testosterone daily. The **seminiferous tubules** comprise the bulk of the testes and are responsible for the production of approximately 30 million spermatozoa per day during male reproductive life (puberty to death). The Sertoli cells, the nurse cells of the seminiferous tubules, are the other major endocrine cells, and they produce inhibin B, anti-Müllerian hormone (AMH) and products important for regulating spermatogenesis.

The two testicular components are interrelated, and both require an intact hypothalamic-pituitary axis for initiation and maintenance of their function. In addition, accessory genital structures (including the epidydimi) are required for the functional maturation and transport of spermatozoa. Thus, disorders of the testes, hypothalamus, pituitary, or accessory structures may result in abnormalities of androgen or gamete production, infertility, or a combination of these problems.

ANATOMY AND STRUCTURE-FUNCTION RELATIONSHIPS (FIGURE 12–1)

TESTES

The adult testis is a prolate spheroid with a normal volume that ranges from 15 to 30 cc. The mean volume is about 20 cc. The average length measures 3.5 to 5.5 cm, and the average width is 2.0 to 3.0 cm. The testes are located within the scrotum, which not only serves as a protective envelope but also helps to maintain the testicular temperature approximately 2°C below the body core temperature. Three layers of membranes—visceral tunica vaginalis, tunica albuginea, and tunica vascular—comprise the testicular capsule. Extensions of the tunica albuginea into the testicle as fibrous septa result in the formation of approximately 250 to 300 pyramidal lobules, each of which contains coiled seminiferous tubules. Within each testis there are over 400 meters of seminiferous tubules, and these structures account for about half of the testicular mass. The approximately 200 million androgen-producing Leydig cells, as well as the blood and lymphatic vessels, nerves, and fibroblasts, are interspersed between the seminiferous tubules.
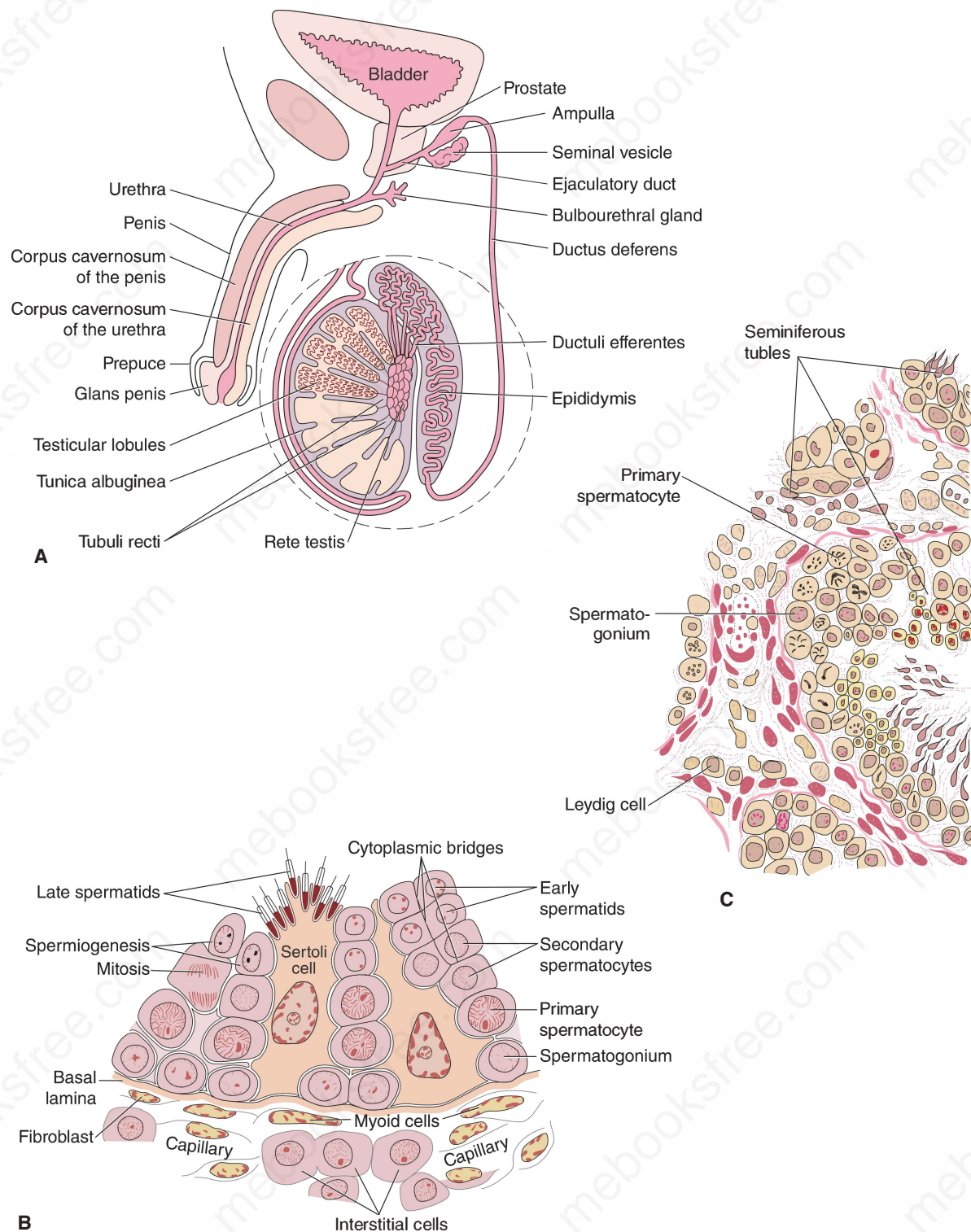
FIGURE 12–1 Male genital system. **A.** The testis and the epididymis are in different scales from the other parts of the reproductive system. Observe the communication between the testicular lobules. **B.** Structural organization of the human seminiferous tubule and interstitial tissue. This figure does not show the lymphatic vessels frequently found in the connective tissue. (**A** and **B** reproduced with permission from Junqueira LC, Carneiro J, Kelley RO. *Basic Histology*. 9th ed. New York: McGraw-Hill Education; 1999.) **C.** Section of human testis. (**C** reproduced with permission from Ganong WF. *Review of Medical Physiology*. 20th ed. New York: McGraw-Hill Education; 2001.)

The blood supply to the testes is derived chiefly from the testicular arteries that are branches of the internal spermatic arteries. After traversing a complicated capillary network, blood enters multiple testicular veins that form an anastomotic network, the lamniform plexus. The lamniform plexuses coalesce to form the internal spermatic veins. The right spermatic vein drains directly into the vena cava; the left enters the renal vein.

The seminiferous tubules in the adult average 165 μm in diameter and are composed of Sertoli cells and germinal cells. The Sertoli cells line the basement membrane and form tight junctions with other Sertoli cells. These tight junctions prevent the passage of proteins from the interstitial space into the lumens of the seminiferous tubules, thus establishing a *blood–testis barrier*. Through extension of cytoplasmic processes, the Sertoli cells surround developing germ cells and provide an environment essential for germ cell differentiation. In addition, Sertoli cells regulate the movement of differentiating germ cells from the base of the tubule toward the lumen and the release of mature sperm into the seminiferous tubular lumen. Sertoli cells also actively phagocytose damaged germ cells and residual bodies that are portions of the germ cell cytoplasm not used in the formation of spermatozoa. Sertoli cells secrete androgen-binding protein, a molecule with high affinity for androgens, into the seminiferous tubular lumen. This protein helps to provide a high concentration of testosterone to the developing germinal cells during the process of spermatogenesis; intratesticular testosterone concentrations are 60 to 100 fold higher than circulating concentrations. Finally, Sertoli cells secrete inhibin B, a protein that is an important regulator of pituitary follicle-stimulating hormone (FSH) secretion.

More than a dozen different types of germ cells have been described in males. Broadly, they can be classified as spermatogonia, primary spermatocytes, secondary spermatocytes, spermatids, and spermatozoa. Spermatogenesis occurs in an orderly fashion, with the spermatocytes being derived from the spermatogonia via mitotic division. Through meiotic (or reduction) division, the spermatids are formed; they contain a haploid number of chromosomes (23). The interval from the beginning of spermatogenesis to release of mature spermatozoa into the tubular lumen is 74 (±4) days. Although there is little variation in the duration of the spermatogenic cycle, a cross-section of a seminiferous tubule demonstrates several stages of germ cell development.

ACCESSORY STRUCTURES

The seminiferous tubules empty into a highly convoluted anastomotic network of ducts called the **rete testis**. Spermatozoa are then transported through efferent ductules into a single duct, the epididymis, by testicular fluid pressure, ciliary motion, and contraction of the efferent ductules. During the 12 to 21 days required for transit from the seminiferous tubular lumen through the epididymis, spermatozoa undergo morphologic and functional changes essential to confer on the gametes the capacity for fertilizing an ovum. The epididymis also serves as a reservoir for sperm. Spermatozoa stored in the epididymis enter the vas deferens, a

muscular duct 35 to 50 cm long that propels its contents by peristaltic motion into the ejaculatory duct.

In addition to the spermatozoa and the secretory products of the testes, retia testis, and epididymi, the ejaculatory ducts receive fluid from the seminal vesicles. These paired structures, 4.5 to 5.5 cm long, are composed of alveolar glands, connective tissue, and muscle. They are the source of seminal plasma fructose, which provides nourishment to the spermatozoa. In addition, the seminal vesicles secrete seminogelin I, phosphorylcholine, ergothioneine, flavins, prostaglandins, and antioxidants, including superoxide dismutase, catalase, and ascorbic acid. About 60% of the total volume of seminal fluid is derived from the seminal vesicles.

The ejaculatory ducts terminate in the prostatic urethra. There additional fluid (~20% of total volume) is added by the prostate, a tubuloalveolar gland with a fibromuscular stroma that weighs about 20 g and measures 4 cm × 2 cm × 3 cm. The constituents of the prostate fluid include spermine, citric acid, cholesterol, phospholipids, fibrinolysin, fibrinogenase, zinc, acid phosphatase, and prostate-specific antigen, a 34-kDa kallikrein-like serine protease. Fluid is also added to the seminal plasma by the bulbourethral (Cowper) glands and urethral (Littre) glands during its transit through the penile urethra.

PHYSIOLOGY OF THE MALE REPRODUCTIVE SYSTEM

GONADAL STEROIDS (FIGURE 12–2)

The three steroids of primary importance in male reproductive function are testosterone, dihydrotestosterone (DHT), and estradiol. From a quantitative standpoint, the most important androgen is testosterone. Testicular Leydig cells produce over 95% of circulating testosterone. In addition to testosterone, the testes secrete small amounts of the potent androgen DHT and the weak androgens dehydroepiandrosterone (DHEA) and androstenedione. The Leydig cells also secrete small quantities of estradiol, estrone, pregnenolone, progesterone, 17α-hydroxypregnenolone, and 17α-hydroxyprogesterone. The steps in testicular androgen biosynthesis are illustrated in Figure 12–2.

DHT and estradiol are also produced by target tissue conversion of testicular and adrenal-secreted androgen and estrogen precursors. Thus, about 80% of the circulating concentration of these two steroids is derived from peripheral conversion. Table 12–1 summarizes the approximate contributions of the testes, adrenals, and peripheral tissues to the circulating levels of several sex steroid hormones in men.

In the blood, androgens and estrogens exist in either a free (unbound) state or as bound to serum proteins. Although about 38% of testosterone is weakly bound to albumin, the major binding protein is *sex hormone–binding globulin* (SHBG), a high affinity, low capacity binding protein that binds 60% of the testosterone. This glycosylated dimeric protein is homologous to the androgen-binding protein secreted by the Sertoli cells. SHBG is synthesized in the liver, with the gene located on the short arm of chromosome 17.
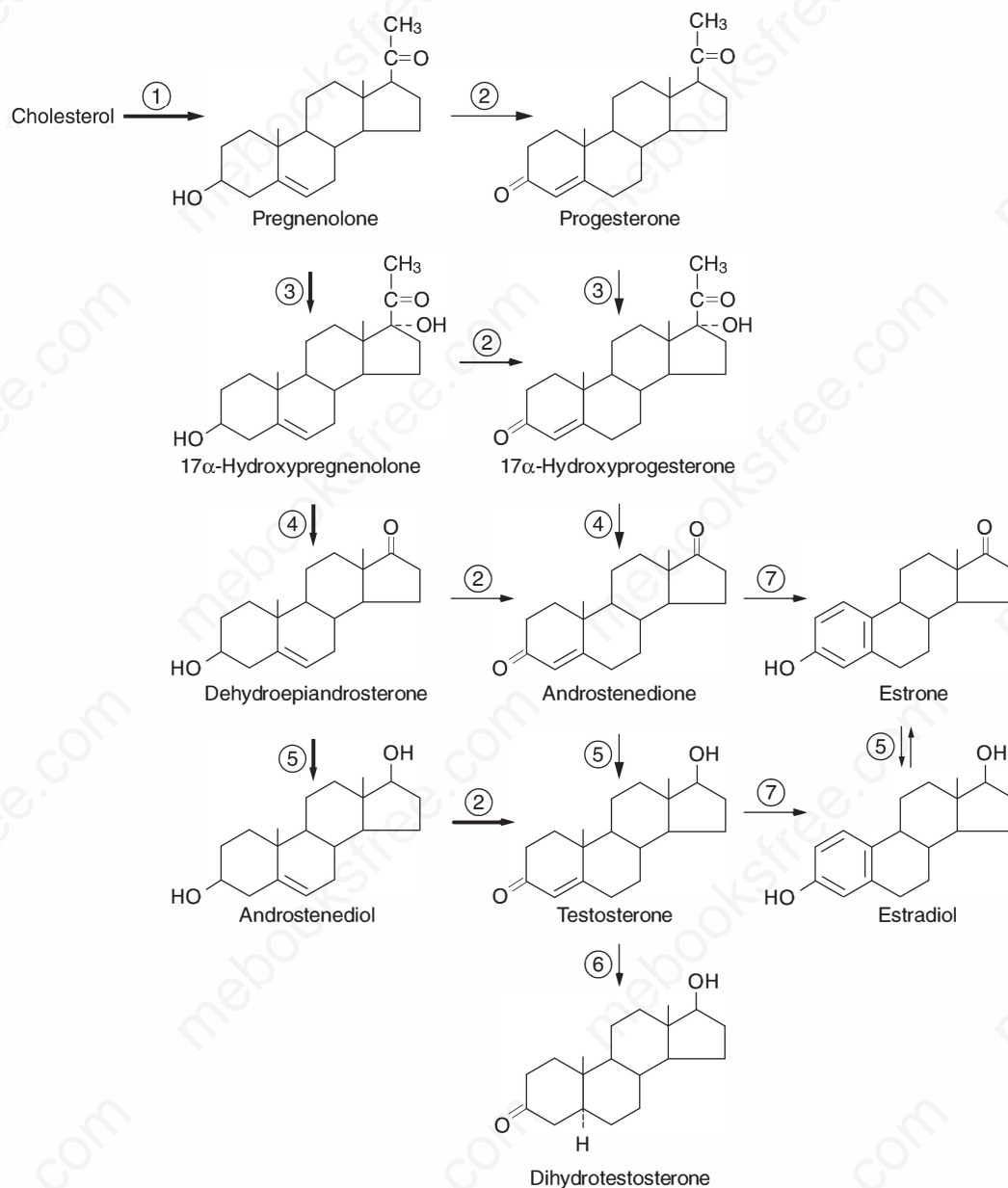
FIGURE 12–2 Pathways for testicular androgen and estrogen biosynthesis. Heavy arrows indicate major pathways. Circled numbers represent enzymes as follows: ①, CYP11A1 (20,22-desmolase; P450scc); ②, 3β-hydroxysteroid dehydrogenase and Δ^5,Δ^4-isomerase II (3βHSDII); ③, CYP17 (17α-hydroxylase; P450c17); ④, CYP17 (17,20-lyase; P450c17); ⑤, 17α-HSD3 (17α-hydroxysteroid dehydrogenase 3; 17-ketoreductase); ⑥, 5α-reductase; ⑦, CYP19 (aromatase). StAR, steroidogenic acute regulatory protein. (See also Figures 9–4, 13–4, and 14–13.)

Certain drugs (including estrogen, tamoxifen, many anti-epileptic drugs, exogenous thyroid hormone) and diseases (including untreated hyperthyroidism) increase serum SHBG concentrations while exogenous androgen, glucocorticoid, or growth hormone therapy and untreated hypothyroidism, obesity, and hyperinsulinemic states decrease serum SHBG concentrations.

About 2% of the circulating testosterone is not bound to serum proteins. According to the free testosterone hypothesis, unbound or "free" testosterone is the bioactive hormone. According to this hypothesis, testosterone is avidly bound to SHBG and is not available to tissues. Testosterone is weakly bound to albumin, and it

may dissociate rapidly from testosterone to enter target tissues. Testosterone bound to albumin is often described as "weakly bound" testosterone and is considered to be bioavailable to target tissues. The sum of unbound, free testosterone plus weakly bound testosterone is often called "bioavailable testosterone." There is relatively little human data to support the free testosterone hypothesis, but it remains the most prevalent theory to explain the effects of circulating testosterone.

Testosterone is converted to DHT and estrogen within specific androgen target tissues. Most circulating testosterone is converted primarily by the liver into various metabolites such as

TABLE 12–1 Relative contributions (approximate percentages) of the testes, adrenals, and peripheral tissues to circulating levels of sex steroids in normal men.

Hormone	Testicular Secretion	Adrenal Secretion	Peripheral Conversion of Precursors
Testosterone	95	<1	<5
Dihydrotestosterone	<20	<1	80
Estradiol	<20	<1	80
Estrone	<2	<1	98
DHEA sulfate	<10	90	—

androsterone and etiocholanolone, which, after conjugation with glucuronic or sulfuric acid, are excreted in the urine as 17-ketosteroids. However, it should be noted that only 20% to 30% of the urinary 17-ketosteroids are derived from testosterone metabolism. The majority of the 17-ketosteroids are formed from the metabolism of adrenal steroids. Therefore, 17-ketosteroid determinations do not reliably reflect testicular steroid secretion.

Testosterone leaves the circulation and rapidly traverses the cell membrane (Figure 12–3). In most androgen target tissues, including the male urogenital tract, testosterone is enzymatically converted to the more potent androgen DHT by the microsomal isoenzyme 5α-reductase-2, which has a pH optimum of 5.5. In the skin, another isoenzyme, 5α-reductase-1 which has a pH optimum near 8.0, converts testosterone to DHT. 5α-reductase-1

FIGURE 12–3 Mechanisms of androgen action (DHT, dihydrotestosterone; mRNA, messenger RNA; Rc, inactive receptor; Rn, activated nuclear receptor; T, testosterone).

activity contributes to male pattern baldness. This enzyme is not found in the urogenital tract.

DHT and testosterone bind to the same specific intracellular receptor protein (R_c in Figure 12–3) that is distinct from both androgen-binding protein and SHBG. The genes that encode for the androgen receptor are located on the X chromosome. The androgen receptor is a member of the steroid-thyroid hormone nuclear receptor superfamily. It is synthesized in the cytoplasm, and it is associated with several heat shock proteins and immunophilin chaperones that modulate receptor structure. When testosterone or DHT binds to the androgen receptor, the androgen–androgen receptor complex dissociates from the heat shock and chaperone proteins, and conformational changes take place that allow it to form a homodimer with another androgen–androgen receptor molecule and to be translocated into the nucleus through binding to transporter proteins, importins (R_n in Figure 12–3). In the nucleus, the androgen–androgen receptor complex binds to androgen response elements in DNA and interacts with protein coactivators to initiate transcriptional activity. Thus, the classic pathway of androgen action results from changes in protein synthesis, a process that takes hours to days. In addition to this classical genomic pathway, androgens may exhibit a rapid effect (seconds to minutes) through nonclassical signaling pathways via cell surface actions that alter traditional secondary messengers such as calcium, cyclic AMP, and others.

A variety of biologic effects of androgens have been defined in humans. As discussed in Chapter 14, they are essential for appropriate differentiation of the internal and external male genital system during fetal development. During puberty, androgen-mediated growth of the scrotum, epididymis, vas deferens, seminal vesicles, prostate, and penis occurs. The normal development of these organs requires androgens. Androgens stimulate skeletal muscle growth and growth of the larynx, which results in deepening of the voice, and of the epiphyseal cartilaginous plates, which results in the pubertal growth spurt. Both ambisexual (pubic and axillary) hair growth and sexual (beard, mustache, chest, abdomen, and back) hair growth are stimulated, as is sebaceous gland activity. Other effects include stimulation of erythropoiesis and social and behavioral changes.

CONTROL OF TESTICULAR FUNCTION

Hypothalamic-Pituitary-Leydig Cell Axis (Figure 12–4)

The hypothalamus synthesizes gonadotropin-releasing hormone (GnRH), a decapeptide, and secretes it in pulses every 30 to 120 minutes into the hypothalamohypophyseal portal blood. After reaching the anterior pituitary, GnRH binds to the receptors on gonadotrophs and stimulates the pulsatile release of luteinizing hormone (LH) and, to a lesser extent, FSH into the general circulation. LH binds specific membrane receptors on Leydig cells. The LH receptor is a G protein–coupled receptor containing seven transmembrane domains with a serine- and threonine-rich cytoplasmic region containing a phosphorylation site and a 350 to 400
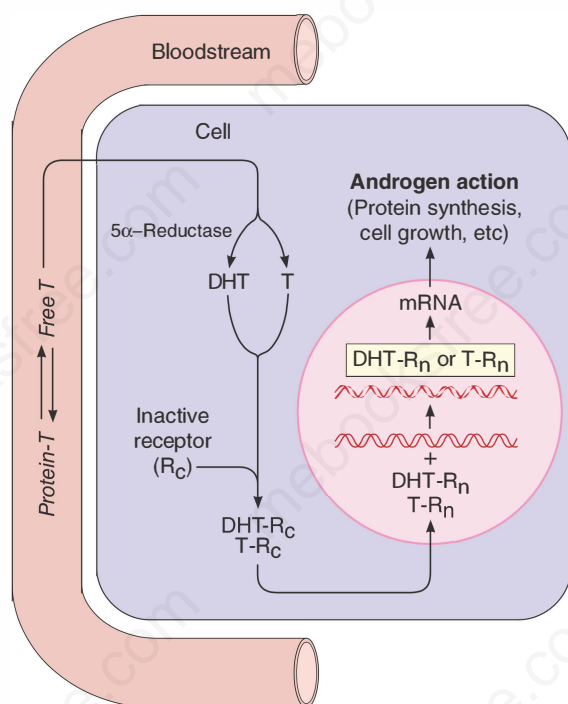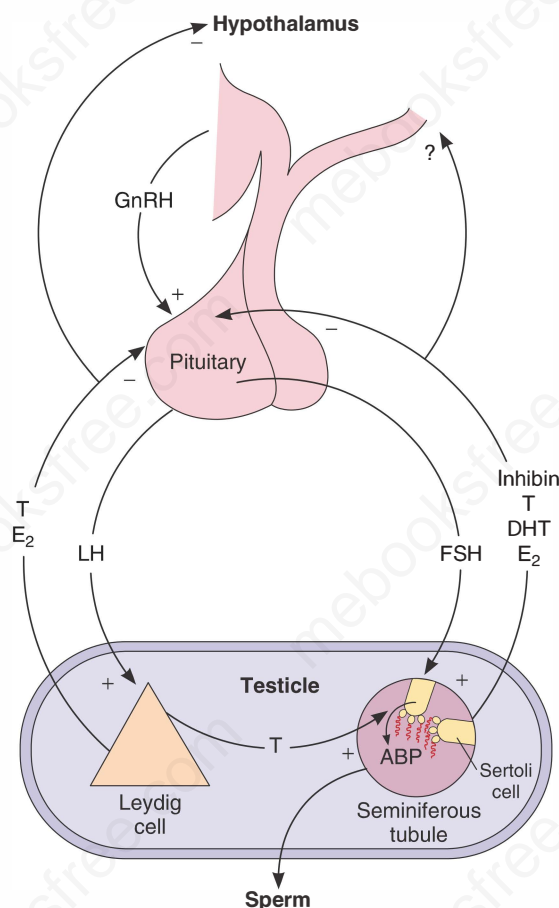
FIGURE 12–4 Hypothalamic-pituitary-testicular axis (ABP, androgen-binding protein; DHT, dihydrotestosterone; E$_2$, estradiol; FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; T, testosterone; +, positive influence; –, negative influence).

amino acid extracellular hormone-binding domain. The binding of LH to its receptor leads to activation of adenylyl cyclase and generation of cAMP and other messengers. It also stimulates the synthesis of steroidogenic acute regulatory (StAR) protein. StAR enhances the transfer of cholesterol into the inner mitochondrial membrane of Leydig cells, leading to increased synthesis and secretion of sex steroid hormones.

The hypothalamus and the pituitary have androgen and estrogen receptors. Experimentally, pure androgens such as DHT reduce LH pulse frequency, whereas estradiol reduces LH pulse amplitude. However, the major inhibitory effect of androgen on the pituitary hypothalamus is mediated by estradiol that is derived by local aromatization of testosterone. Leydig cells also secrete small quantities of cytokines and other peptides that are important for paracrine regulation of testicular function.

Hypothalamic-Pituitary-Seminiferous Tubular Axis (See Figure 12–4)

After stimulation by GnRH, the gonadotrophs secrete FSH into the systemic circulation. This glycoprotein hormone binds to specific receptors in the Sertoli cells and stimulates the production of inhibin B and several other factors including insulin-like growth factor-I, transferrin, and Müllerian duct inhibitory factor. FSH is necessary for quantitatively and qualitatively normal spermatogenesis. Its effects on spermatogenesis include maintenance of very high intratesticular testosterone concentrations by stimulation of Sertoli cell production of androgen-binding protein.

Although two forms of inhibin have been identified, inhibin B is the physiologically form in men. This 32-kDa protein, which is composed of an alpha and a beta subunit, selectively inhibits FSH release from the pituitary without affecting LH release. FSH directly stimulates the Sertoli cells to secrete inhibin B. Serum inhibin B concentrations reflect Sertoli cell number and sperm production and, therefore, correlate with testicular size. There is a reciprocal relationship between serum inhibin B and FSH concentrations: inhibin B levels are lower and FSH concentrations are higher in conditions associated with lower levels of spermatogenesis. Inhibin B is the most important physiologic regulator of pituitary FSH secretion, and, together with circulating sex steroids, provides negative feedback on FSH secretion.

Activin and transforming growth factor β (TGFβ), inhibin B-related proteins have been identified in human testes. These factors can selectively stimulate pituitary FSH secretion *in vitro*, but their role in the physiologic regulation of FSH secretion in men is unknown. It is more likely that these proteins have a local paracrine effect in the human testis.

EVALUATION OF MALE GONADAL FUNCTION

CLINICAL EVALUATION

Clinical Presentation

The clinical presentation of patients with deficient testosterone production or action depends on the age at onset of hypogonadism. Androgen deficiency during the second to third months of fetal development results in varying degrees of ambiguity of the genitalia and male pseudohermaphroditism. If the deficiency develops during the third trimester, defects in testicular descent leading to cryptorchidism as well as micropenis may occur. These topics are also discussed in Chapter 14.

Postnatal prepubertal androgen deficiency leads to poor secondary sexual development and eunuchoidal skeletal proportions. The penis fails to enlarge, the testes remain small (≤2 cc), and the scrotum does not develop the marked rugae characteristic of puberty. The voice remains high-pitched, and the muscle mass does not develop fully, resulting in less than normal strength and endurance. The lack of appropriate stimulation of sexual hair growth results in sparse axillary and pubic hair (which receive some stimulation from adrenal androgens) and absent or very sparse facial, chest, upper abdominal, and back hair. Although the androgen-mediated pubertal growth spurt fails to take place, the epiphyseal plates of the long bones continue to grow under the
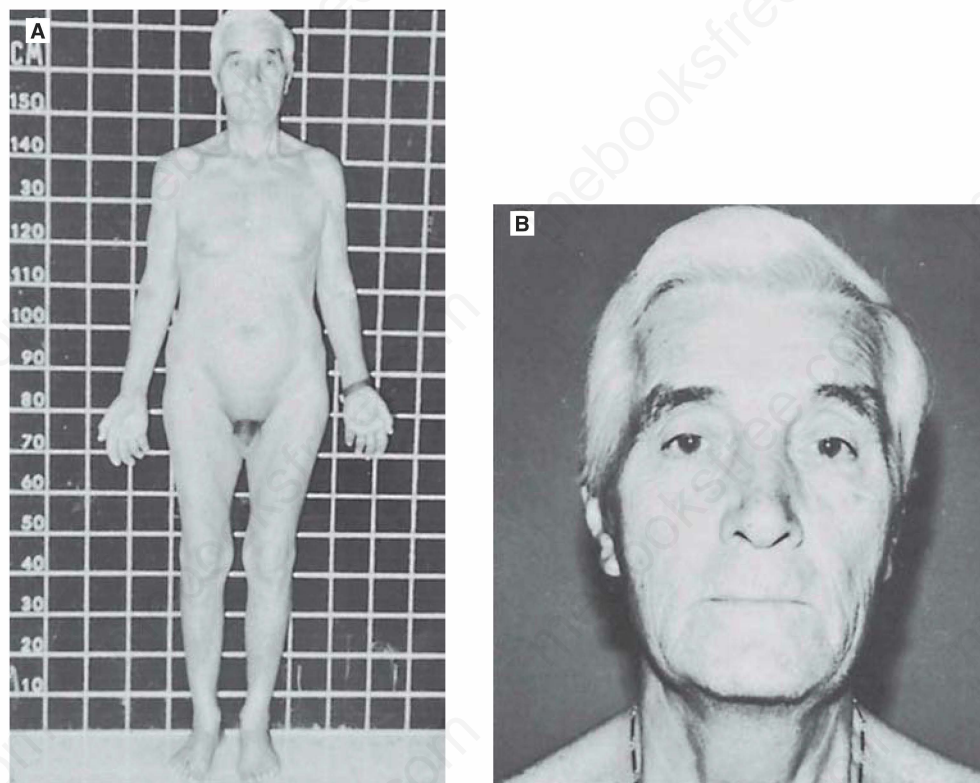
FIGURE 12–5   **A.** Hypogonadal habitus. Note absence of body and facial hair as well as feminine body distribution. **B.** Hypogonadal facies. Note absence of facial hair and fine wrinkles around the corners of the eyes and lips.

influence of insulin-like growth factor-I and other growth factors. Thus, the long bones of the upper and lower extremities grow out of proportion to the axial skeleton. Healthy white men have an average upper segment (crown to pubis) to lower segment (pubis to floor) ratio of greater than one, whereas prepubertal hypogonadism results in a ratio of less than one. Similarly, the ratio of total arm span to total height averages 0.96 in white men. Because of the relatively greater growth in the upper extremities, the arm span of eunuchoidal individuals exceeds height by 5 cm or more.

If testosterone deficiency develops after puberty, the patient may complain of decreased libido, erectile dysfunction, and low energy. Patients with mild androgen deficiency or recent onset of androgen deficiency do not typically present with decreased facial or body hair growth. Although adult androgen levels must be achieved to *stimulate* male sexual hair growth, relatively low levels of androgens are required to *maintain* sexual hair growth. After years of hypogonadism, the growth of facial hair may diminish, and the frequency of shaving may decrease (Figure 12–5). In addition, fine wrinkles may appear in the corners of the mouth and eyes.

Genital Examination

Adequate assessment of the genitalia is essential in the evaluation of male gonadal disorders. The examination should be performed in a warm room in order to relax the dartos muscle of the scrotum. If necessary, a warm towel may be applied to the scrotum for a few minutes to further relax the dartos and allow a retracted testis to

drop into the scrotal sac. The penis should be examined for the presence of hypospadias, epispadias, and chordee (abnormal angulation of the penis due to a fibrotic plaque) that might interfere with erectile function and fertility. If clinically indicated, the fully stretched dorsal penile length should be measured in the flaccid state from the pubopenile skin junction to the tip of the glans. The normal range in adults is 12 to 16 cm (10th and 90th percentiles, respectively).

Accurate measurement of testicular volumes is important. Historically, careful measurement with calipers of the longitudinal and transverse axes of the testes could be used to estimate testicular volume (V) with the formula for a prolate spheroid: $V = 0.52 \times \text{length} \times \text{width}^2$. However, this method is uncomfortable for the patient and is subject to significant variability. A Prader orchidometer is a more reproducibly accurate method. The Prader orchidometer consists of a series of plastic ellipsoids ranging in volume from 1 to 35 mL. Each testis is compared with the appropriate ellipsoid. The epididymis should not be included in the volume estimation. Adults normally have a testicular volume more than or equal to 15 mL by this method. Prader orchidometry correlates well with volume determined by ultrasound although the volumes measured by ultrasound are consistently smaller. A normal testicular volume measured by ultrasound is more than 10 mL.

Because about the majority of human testicular volume (>80%) is composed of seminiferous tubules, low testicular volumes correlate with decreased spermatogenesis due to lack of tubular development (a prepubertal disorder) or regression of tubular size (a postpubertal disorder). The consistency of the testes

should be noted. Small, firm testes are characteristic of hyalinization or fibrosis, a finding suggestive of Klinefelter syndrome. Men with postpubertal testicular atrophy may have soft testes of low-normal or normal volumes.

The epididymis and vas deferens should also be examined. Acute epididymitis is characterized by marked epididymal tenderness. A firm or enlarged epididymis suggests chronic epididymitis. The normal vas deferens feels like a small firm cord or string in the larger, spongy spermatic cord. The absence of one or both vasa deferens in an infertile man is highly suggestive of cystic fibrosis; congenital absence of the vas(a) deferens may be the only manifestation of cystic fibrosis. In an infertile man, examination for a large varicocele is important. The patient should be examined in the upright position while performing the Valsalva maneuver. About 85% of varicoceles are located on the left side, and 15% are bilateral.

LABORATORY TESTS OF TESTICULAR FUNCTION

Serum Testosterone Measurement

In normal healthy young and old men, serum testosterone concentrations peak in the early morning and decline to a nadir in late afternoon and night. The normal range for testosterone is based on measurements done on early morning samples from men with normal reproductive function. A normal serum testosterone concentration measured on an afternoon or evening serum sample virtually excludes male hypogonadism.

Because reliable total testosterone assays are generally widely available and inexpensive, the Endocrine Society guidelines and most experts advocate measurement of total testosterone as the initial serum testosterone test. A total testosterone concentration that is clearly in the normal range (>400 ng/dL) generally excludes hypogonadism when measured using a reliable total testosterone assay with a lower limit of normal of 275 to 300 ng/dL.

When the clinician suspects that the patient might have an abnormality in SHBG concentrations (eg, patients with diabetes mellitus, obesity, or hepatopathy and patients taking corticosteroids, anti-epileptics or age ≥65), many experts would recommend measurement of serum free testosterone by an accurate method. Men with Klinefelter syndrome may have relatively high serum estradiol concentrations that will raise SHBG concentrations. For men with suspected Klinefelter syndrome, it is prudent to measure a free testosterone concentration, even if the serum total testosterone concentration is well within the normal range. For many endocrinologists, the conditions that affect serum SHBG concentrations (Table 12–2) are so common that measurement of serum free testosterone might be useful in the majority of their patients. Measurement of serum free testosterone concentrations by an accurate method is also useful in men with borderline low or low-normal serum total concentrations.

The most commonly available total testosterone assays are immunoassays. These assays are convenient and inexpensive. They are generally reliable although the clinician must confirm whether

TABLE 12–2 Conditions commonly associated with altered serum SHBG concentrations.

Lower SHBG	Higher SHBG
Obesity	Hepatopathy
Corticosteroids	Anti-epileptics
Hypothyroidism	Hyperthyroidism
Androgens	Estrogens

an appropriate cohort was used to establish the normal range. The normal range should be determined using early morning blood samples from 60 to 100 healthy young men—ideally with proven fertility. Laboratories often use the normal range supplied by the manufacturer or establish a normal range on an inappropriate convenience sample such as those from blood donors. Immunoassays tend to be less accurate and reliable in samples with very low or very high serum total testosterone concentrations. The most reliable methods are specific immunoassays following extraction of serum or gas chromatography or liquid chromatography-tandem mass spectroscopy. Such assays require expensive equipment, but they are becoming more common and offer significant advantages over traditional immunoassays.

When clinically indicated or for a patient with borderline low or low-normal serum total testosterone concentration, free testosterone concentration should be determined. The gold standard method for measurement of free testosterone is equilibrium dialysis, but this method is expensive, slow, and not widely available. Serum free testosterone concentrations can be calculated based on the measurement of total testosterone, SHBG, and albumin concentrations. There are several validated formulae for the calculation of free testosterone concentrations. There is some controversy about which method is the best, but several of the formulae provide results that correlate highly with values measured by equilibrium dialysis. Platform serum free testosterone assays that use an analog methodology are inaccurate and significantly underestimate true serum free testosterone concentrations. These assays are still widely available, but they should not be used.

At least as important as the choice of the testosterone assay is to understand the principles of when to test and how to interpret the test results. Any acute systemic or flare of chronic illness may significantly suppress the gonadal axis resulting in a low serum total and free testosterone concentrations. Laboratory assessment for male hypogonadism should occur when a man is not acutely ill. In addition, there is significant variation in day-to-day serum testosterone concentrations (10%-20%). Before making the diagnosis of male hypogonadism, serum testosterone concentrations must be demonstrated to be low on at least two occasions on blood samples obtained in the early morning. In addition, measurement of serum testosterone during the fasting state reduces the variability and results in higher serum testosterone concentrations. Therefore, measurement of testosterone on an early morning blood sample obtained after 6 to 8 hours of fasting may be useful in many men with equivocal testosterone concentrations.

Serum Estradiol Measurement

Serum estradiol concentrations are at or below the lower limit of sensitivity for most commercial immunoassays. It is rarely necessary or useful to measure serum estradiol concentrations in men. However, when clinically indicated, a clinician may send a serum sample to a specialty laboratory that uses liquid chromatography-tandem mass spectroscopy to measure serum estradiol concentrations.

Gonadotropin and Prolactin Measurements

LH and, to a lesser extent, FSH are released in pulsatile fashion throughout the day. In addition, there is a diurnal variation with higher levels in the morning than in the evening. In general, serum FSH and LH concentrations should be measured simultaneously on a single serum sample obtained between 7 and 10 AM. As a strategy to address the variability of serum testosterone, FSH and LH concentrations, some experts recommend obtaining at least three blood samples at 20- to 40-minute intervals, and measuring testosterone, FSH, LH concentrations in a single pooled specimen. In practice, pooled samples are seldom necessary.

The concentrations of gonadotropins measured in one laboratory are generally not directly comparable to those measured in another because of differences in assays, reference preparations, cohorts used to establish normal ranges, and other factors. The lower limit of normal for an accurate assay should be less than 1.0 IU/L. The upper limit of normal serum FSH is 7 IU/L in men of confirmed normal reproductive and gonadal function; laboratories that report a higher normal range for FSH are likely not using men of normal reproductive capacity to establish the normal range.

The primary use of basal FSH and LH concentrations is to distinguish between primary (hypergonadotropic) hypogonadism (both gonadotropins are elevated), and secondary (hypogonadotropic) hypogonadism, in which the gonadotropins are low or inappropriately normal in the presence of decreased androgen production. In primary hypogonadism, serum FSH concentrations are always higher than LH concentrations because the decreased negative feedback from inhibin B and sex steroids has greater effects on FSH secretion than the decreased negative feedback of sex steroids on LH secretion. In secondary hypogonadism, the pituitary may secrete gonadotropins that are immunoreactive, but not bioactive, thus resulting in inappropriately normal serum gonadotropin concentrations.

Elevations of serum prolactin (PRL) inhibit the normal release of pituitary gonadotropins (shown by a reduced LH pulse frequency), probably through an effect on the hypothalamus.

Serum PRL measurements should be performed in any patient with hypogonadotropic hypogonadism. Serum PRL concentrations are generally stable throughout the day; therefore, measurement of this hormone in a single sample is usually sufficient. However, the patient should abstain from eating for 3 hours before the blood sample is obtained, because a protein meal may acutely stimulate the release of PRL from the pituitary.

TABLE 12–3 Typical normal ranges for sex steroids, gonadotropins, and prolactin in normal men.*

Hormone	Ranges
Testosterone, total	280-900 ng/dL (10.0-31.3 nmol/L)
Testosterone, free	50-210 pg/mL (173-729 pmol/L)
Dihydrotestosterone	27-75 ng/dL (0.9-2.6 nmol/L)
Androstenedione	50-250 ng/dL (1.7-8.5 nmol/L)
Estradiol	10-50 pg/mL (3.67-18.35 pmol/L)
Estrone	15-65 pg/mL (55.5-240 pmol/L)
FSH	1.0-7.0 mIU/mL (1.0-7.0 IU/L)
LH	1.0-7.0 mIU/mL (1.0-7.0 IU/L)
Prolactin	5-18 ng/mL (218-780 nmol/L)

*Normal ranges will vary modestly based on assay techniques. The normal range should be determined in a group of ≥60 healthy young men with proven normal reproductive function.

Table 12–3 shows the typical normal ranges for serum concentrations of sex steroid hormones, gonadotropins and prolactin in normal healthy men.

In men less than or equal to 50 years old with secondary hypogonadism, iron overload due to hemochromatosis should also be excluded. An iron saturation or ferritin measurement in a single serum sample is usually adequate. In men greater than 50 years old, it is very rare to present with isolated hypogonadotropic hypogonadism as the sole manifestation of hemochromatosis. In men greater than 50 years old, measurement of iron saturation or ferritin is not necessary unless there is clinical suspicion of an iron overload syndrome.

SPECIAL TESTS

Semen Analysis

Seminal fluid analysis is the most important test in the evaluation of male infertility. With some exceptions, a normal semen analysis excludes significant gonadal dysfunction. However, a single abnormal semen analysis is not a sufficient basis for a diagnosis of disturbance of testicular function, because marked variations in several of the parameters may be seen in normal individuals: at least three semen samples should be examined over a 2- to 3-month interval in order to evaluate this facet of male gonadal function. As noted earlier, approximately 3 months are required for completion of the spermatogenic cycle and movement of the mature spermatozoa through the ductal system. Therefore, when an abnormal semen sample is produced, one must question the patient about prior fever, trauma, drug exposure, and other factors that may temporarily impair spermatogenesis.

The semen should be collected by masturbation after 2 to 7 days of ejaculatory abstinence and examined within 1 hour after collection. Normal semen has a volume of more than 2 to 4 mL, with 15×10^6 or more sperms per milliliter. Over half of the spermatozoa should exhibit progressive motility, and 30% or more should have normal morphology.

Chorionic Gonadotropin Stimulation Test

Human chorionic gonadotropin (hCG) is a glycoprotein hormone with biologic actions similar to those of LH. Following an injection of hCG, this hormone binds to the LH receptors on the Leydig cells and stimulates the synthesis and secretion of testicular steroids. Administration of hCG may be useful to confirm the presence of testicular tissue in the evaluation of cryptorchidism in prepubertal boys. There are a number of published protocols, but one protocol is to administer of 1500 IU hCG twice weekly for 3 weeks. Anorchic boys have no significant increase of serum testosterone following administration of hCG whereas serum testosterone concentrations increase into the normal adult male range in boys with cryptorchidism.

Testicular Biopsy

Testicular biopsy is primarily indicated as a diagnostic study for infertile men with normal-sized testes and azoospermia in order to distinguish between spermatogenic failure and ductal obstruction. Although germinal aplasia, hypoplasia, maturation arrest, and other abnormalities of spermatogenesis may be diagnosed by examination of testicular tissue in oligospermic males, knowledge of the type of defect does not alter therapy. Therefore, testicular biopsy is not usually indicated for evaluation of mild to moderate oligospermia. For treatment of male infertility, testicular biopsies are also done to perform testicular sperm extraction for possible intracytoplasmic sperm injection (ICSI) into ova.

Evaluation for Male Hypogonadism

The prevalence of male hypogonadism is not high enough to justify screening. Even in older men, where the prevalence increases with age, the prevalence appears to be less than 5%. Case finding of hypogonadism should be based on symptoms such as decreased libido and sexual function and signs such as osteoporosis, unexplained anemia, and weakness. Testing for male hypogonadism should be performed when the man is at baseline health and not during acute systemic illness or flares of chronic illness. The initial testing would ideally be done on serum obtained between 7 AM and 10 AM. Initial testing consists of measurement of total testosterone using an accurate assay with a normal range of approximately 300 to 900 ng/dL. The authors advocate for simultaneous measurement of FSH and LH on the initial sample because the results can be helpful for interpreting the serum testosterone concentration, particularly when the testosterone concentration is in the low-normal range. Assessment of free serum testosterone concentration is useful in men with borderline low to low-normal serum concentrations. Low serum testosterone concentrations must be confirmed with at least one additional measurement from an early morning blood sample (after the patient has fasted for at least 6 hours).

Serum gonadotropin concentrations are useful to determine the etiology of male hypogonadism and to determine additional diagnostic evaluation. Elevated gonadotropin concentrations indicate primary hypogonadism due to testicular dysfunction. Serum karyotyping is useful to diagnose Klinefelter syndrome although this diagnosis may often be made clinically in males over age 16

with very small testes and biochemical evidence of primary hypogonadism. Men with symptoms and signs of hypogonadism, unequivocally low serum testosterone concentrations and inappropriately normal or low serum gonadotropins have secondary hypogonadism due to hypothalamic or pituitary disease. Common causes of secondary hypogonadism include Cushing syndrome, hyperprolactinemia, and sellar tumors or infiltrative disease. In younger men, iron overload syndromes such hereditary hemochromatosis should also be considered. For men with secondary hypogonadism, evaluation for Cushing syndrome consists of review of prescription and over-the-counter medications and supplements, querying for common symptoms of hypercortisolism (weakness, easy bruisability) and physical examination. All men with secondary hypogonadism should have a serum prolactin. Although some experts recommend sella imaging in the evaluation of secondary hypogonadism only in those men with hyperprolactinemia, the authors recommend sella imaging in all men less than or equal to age 50 (even in those with normoprolactinemia) to exclude a pituitary tumor or mass. Imaging by computed tomography is sufficient to exclude a large sellar mass that could cause gonadotropin deficiency. Magnetic resonance imaging is more commonly ordered, but it is more expensive and will commonly detect pituitary microadenomas that are not the cause of secondary hypogonadism. The yield of sella imaging is very low in men greater than age 50, and the authors recommend sella imaging for men greater than age 50 only for those who have new onset of headaches, visual complaints, signs of cranial neuropathy, serum testosterone less than or equal to 150 ng/dL, low serum FSH and LH concentrations or hyperprolactinemia. Even minor hyperprolactinemia is an indication for sellar imaging in all men with secondary hypogonadism. Figure 12–6 outlines an approach to the diagnosis of male hypogonadism.

DRUGS USED FOR TESTOSTERONE REPLACEMENT THERAPY IN MALE HYPOGONADISM

ANDROGENS

A variety of drugs are available for the treatment of androgen deficiency.

Oral Androgens

Alkylated forms of testosterone, such as methyltestosterone, oxymetholone, and fluoxymesterone can be taken orally, but are associated with hepatopathy and have erratic absorption and decreased effectiveness compared to the intramuscular preparations. Oral testosterone undecanoate has been safely prescribed in many countries for decades, but it has a very short half-life and requires administration two to three times daily. It is not available in the United States. Buccal testosterone tablets have been approved for use in the United States. These tablets are designed
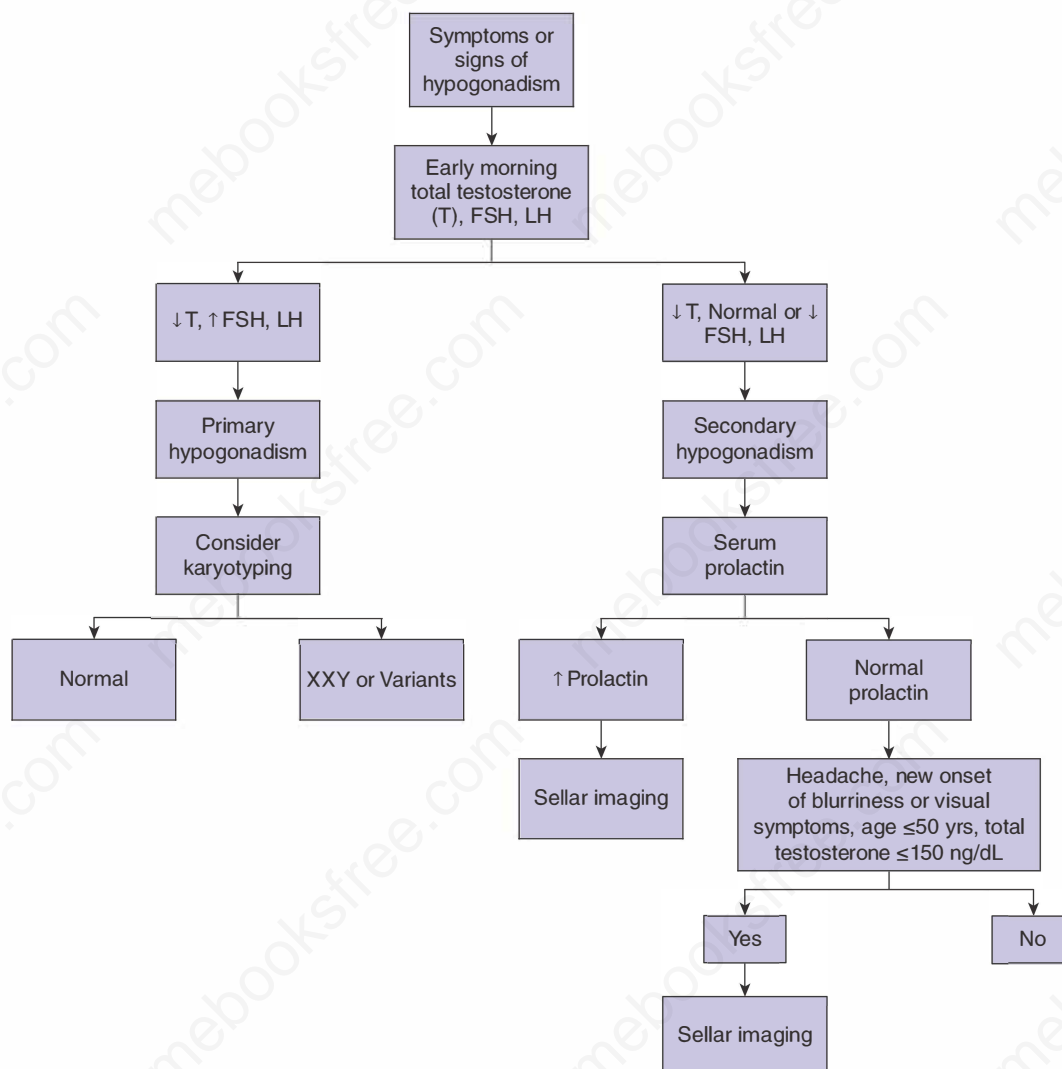
FIGURE 12–6   Algorithm for evaluation of clinical hypogonadism.

to adhere to the gum or inner cheek. This formulation is rarely used because it requires twice daily usage, and some men (5%-10%) experience significant gingival irritation. Novel formulations of oral testosterone and other androgens are under development for treatment of male hypogonadism.

Injectable Testosterone Esters

There are four injectable (intramuscular) testosterone esters that are used clinically: testosterone propionate, enanthate, cypionate, and undecanoate. Testosterone propionate is a short-acting androgen. Its main use is for hypogonadal men with severe bladder outlet obstructive symptoms due to prostate hyperplasia. Obstructive symptoms due to benign prostatic hypertrophy following therapy with testosterone propionate usually resolve rapidly because of its short duration of action. However, because testosterone therapy seldom causes significant worsening of bladder outlet obstruction symptoms, testosterone propionate is very rarely used.

Testosterone enanthate or cypionate have similar pharmacokinetics. The usual dosage of these esters is 75 to 100 mg every week or 150 to 200 mg every 2 weeks. Patients (or relatives) can be taught to self-administer testosterone enanthate and cypionate, and these injectable testosterone esters remain the least expensive, commonly used formulation of testosterone replacement therapy. A long-acting intramuscular formulation of testosterone undecanoate has been approved recently by the Food and Drug Administration (FDA) for use in the United States. The formulation (Aveed) that is available in the United States is administered 750 mg every 10 weeks; some men may need injections every 8 weeks. Clinicians who prescribe Aveed must undergo special training (FDA-approved "Risk evaluation and mitigation strategy"). The FDA has required this training because of the concern of pulmonary oil embolism (from the oily vehicle), a very rare event that usually manifests as cough and transient dyspnea. All Aveed injections must be administered by a healthcare professional in a clinical setting. Outside of the United States, the formulation (Nebido) is dosed 1000 mg every 6 weeks for two dosages followed by 1000 mg every 12 weeks. Injectable testosterone undecanoate has become the most commonly prescribed form of testosterone replacement therapy in many countries outside of

the United States, but it is used much less commonly in the United States because of the inconvenience of clinic visits for administration and expense.

Implantable Testosterone Pellets

Testosterone pellets may be implanted subcutaneously every 3 to 6 months. However, this therapy has not enjoyed much popularity in the United States because it requires a minor office procedure and there is a small risk of extrusion. In addition, the development of injectable testosterone undecanoate has provided a more convenient, long-acting option.

Transdermal Testosterone Therapy

Transdermal delivery of testosterone by patch or gel is the most commonly prescribed testosterone replacement in the United States. Testosterone patches have been intermittently available recently, but a generic testosterone patch is expected to be available in the next few years. There are several different testosterone gels available that are applied daily to (depending on the formulation) the back, flanks, axillae, or thighs. The dosing varies based on the formulation, but the usual starting dosage of testosterone gels is intended to provide 5 to 6 mg of testosterone daily. It is important that the sites of gel application be covered when in close physical contact with children and be cleansed with soap and water prior to heterosexual sexual intercourse in order to avoid the transdermal transfer of the testosterone to children or female partners.

GONADOTROPIN THERAPY

Injectable Human Chorionic Gonadotropin

hCG may be injected subcutaneously or intramuscularly two to three times weekly to induce endogenous testosterone production in men with secondary hypogonadism. Injectable hCG induces aromatization of testosterone to estrogen and is associated with a higher incidence of gynecomastia than other treatments for testosterone deficiency. Because of the need for relatively frequent injections and the higher incidence of gynecomastia with injectable long-term hCG usage, hCG therapy is reserved for treatment of infertility in men with secondary hypogonadism and low levels of spermatogenesis.

Recombinant Human Luteinizing Hormone (rhLH)

Because rhLH is much more expensive than hCG, it has not been used commonly in men.

Side Effects of Testosterone Replacement Therapy

The most common side effects of testosterone replacement therapy are erythrocytosis in older men and acne in younger men. In younger men, significant erythrocytosis seldom occurs whereas

acne rarely occurs in older men. Testosterone induces erythropoiesis by stimulating erythropoietin production and by suppressing hepcidin, an important regulator of iron transport. Testosterone therapy may precipitate male pattern balding in men who are predisposed to that condition, but it reflects normal physiology and is not a true side effect. Transient, tender gynecomastia occurs in a small percentage of men, but it generally resolves within a few weeks of continued treatment. Testosterone therapy tends to suppress serum high density lipoprotein cholesterol concentrations. The effect is more pronounced in younger men with baseline high serum HDL concentrations and with administration of oral alkylated testosterone derivatives. There is no other consistent effect of androgens on serum lipid concentrations.

In older men, testosterone replacement therapy may increase serum prostate specific antigen concentrations, but the effect is usually modest (~0.5-0.6 ng/mL). The 2010 Endocrine Society guidelines recommend that men who experience an increase serum PSA of greater than 1.4 ng/mL in 12 months should be evaluated by a urologist for the possibility of prostate cancer. Androgen therapy currently is contraindicated in patients with prostatic carcinoma, but some clinicians are prescribing testosterone replacement to symptomatic, hypogonadal men with untreated low-grade, localized prostate cancer or intermediate grade, localized prostate cancer that has been successfully treated.

Occasionally, testosterone replacement therapy induces clinically significant sodium and water retention in susceptible individuals. There is significant controversy about whether testosterone replacement therapy is beneficial, neutral or harmful for cardiovascular health.

Another rare side effect of testosterone is worsening or unmasking of underlying sleep apnea. In men who develop erythrocytosis while on modest dosages of testosterone, evaluation for sleep apnea should be done. Although high dosages of synthetic androgenic, anabolic steroids may cause abnormal or aggressive behavior, there is little evidence that physiological testosterone replacement therapy induces aggressive behavior. Hypogonadal men report feeling "irritable," a symptom that improves with testosterone replacement therapy. Some hypogonadal men report feeling more "assertive" while on testosterone replacement therapy.

Monitoring of testosterone replacement therapy includes periodic assessment of hematocrit or serum hemoglobin concentrations in all men. In men treated with transdermal testosterone therapy or men on long-acting intramuscular testosterone undecanoate, it is prudent to check serum testosterone concentrations episodically to assess for adequacy of dosage. For transdermal testosterone, the optimal timing might be 8 to 12 hours after application of the transdermal product, but it is not clear that timing of the blood sample is clinically significant. For intramuscular testosterone undecanoate, a trough serum testosterone should be measured during the week before the next injection. For men using injectable testosterone cypionate or enanthate, routine measurement of serum testosterone is not necessary or useful because of the pharmacokinetics of these intramuscular testosterone esters. With typical dosages of these formulations, testosterone concentrations are supraphysiological at 24 to 72 hours after injection and then fall steadily to the lower limit of normal

by 2 to 3 weeks after the injection. In some men who request higher than usual dosages of injectable testosterone cypionate or enanthate, it is useful to measure serum testosterone concentrations at the mid-interval between injections (when concentrations should be below the upper third of the normal range).

For any formulation of testosterone therapy, the goal is maintenance of serum testosterone concentrations in the normal range, relief of symptoms and signs of hypogonadism and avoidance of adverse effects such as erythrocytosis. Attempting to adjust the dosage of testosterone to the middle of the normal range is not possible because of the variability of testosterone concentrations with any of the currently available therapies. Even with transdermal testosterone formulations that provide the most constant daily testosterone concentrations, targeting a goal testosterone concentration around the middle of the normal range is not feasible. Unfortunately, assessment of serum gonadotropins cannot be used to determine the dosage of testosterone therapy.

For all men on testosterone therapy, the dosage of testosterone should be decreased if erythrocytosis occurs. If erythrocytosis occurs at a low dosage of testosterone, evaluation of sleep apnea should be considered. For all men on testosterone therapy, increasing the dosage of testosterone may be considered if symptoms of hypogonadism have not resolved and erythrocytosis is not present.

It is controversial whether to follow serum prostate specific antigen concentrations in older men using testosterone therapy. There is conflicting evidence about the effects of exogenous testosterone on prostate cancer, and there is also evidence that screening for prostate cancer causes significant harm (from emotional stress) and cost (from procedures including transrectal ultrasound and prostate biopsies) without improving overall mortality. It is the authors' practice to discuss these controversies with older men and to offer them the option of prostate cancer screening with periodic measurements of serum prostate specific antigen.

CLINICAL MALE GONADAL DISORDERS

Hypogonadism may be subdivided into three general categories (Table 12–4). A thorough discussion of the hypothalamic-pituitary disorders that cause hypogonadism is presented in Chapters 4 and 13. The defects in androgen biosynthesis and androgen action are described in Chapter 14. The following section focuses on primary gonadal dysfunction.

SYNDROMES ASSOCIATED WITH PRIMARY GONADAL DYSFUNCTION

Primary hypogonadism is characterized by deficient production of testosterone or its metabolites (estradiol and DHT) and/or defects in spermatogenesis. It has been estimated that 10% to 15% of adult men have isolated defects in spermatogenesis with preservation of Leydig cell function and normal circulating testosterone concentrations. Although there is some controversy about the prevalence of combined hypogonadism with deficient spermatogenesis and

TABLE 12–4 Classification of causes of male hypogonadism.

Hypothalamic-Pituitary Disorders
Panhypopituitarism
Isolated LH and FSH deficiency
  Prepubertal (testes ≤2 cc)
    Congenital
      With hyposmia or anosmia, colobomas, synkinesia (Kallmann syndrome)
      With complex neurologic syndromes
        Prader-Willi syndrome
        CHARGE syndrome
        Laurence-Moon-Bardet-Biedl syndrome
        Cerebellar ataxia syndrome
    Acquired
      Tumor (craniopharyngioma, pituitary adenoma)
      Cushing syndrome
      Hyperporlactinemia
      Severe iron overload syndromes
  Postpubertal (testes >2 cc)
    Congenital with partial LH and FSH deficiency
    Acquired
      Tumor (pituitary adenoma; rarely parasellar tumor)
      Cushing syndrome
      Hyperprolactinemia
      Iron overload syndromes
      Medications (opioids)
      Sleep apnea

Primary Testiscular Abnormalities
Klinefelter syndrome (XXY)
Other chromosomal defects (XX male, XY/XXY, XX/XXY, XXXY, XXXXY)
Bilateral anorchia (vanishing testes syndrome)
Leydig cell aplasia
Cryptorchidism
Noonan syndrome
Myotonic dystrophy
Defects in androgen biosynthesis

Defects in Androgen Action
Complete androgen insensitivity (testicular feminization)
Incomplete androgen insensitivity

CHARGE, Coloboma, Heart defects, Atresia of the nasal choanae, Retardation of growth, Genital abnormalities, Ear defects.

testosterone secretion, it is much less common. The best estimates are that 2% to 4% of middle-aged and older men and 0.5% to 1.0% of younger men have combined hypogonadism.

CAUSES OF PRIMARY HYPOGONADISM PRESENTING IN CHILDHOOD

KLINEFELTER SYNDROME (XXY SEMINIFEROUS TUBULE DYSGENESIS)

Klinefelter syndrome is the most common genetic cause of male hypogonadism, occurring in one of 600 male births. An extra X chromosome is present in about 0.2% of male conceptions and

0.1% to 0.2% of live-born males. Sex chromosome surveys have demonstrated that up to 2.5% of boys and young men with congenital cognitive impairment have an extra X chromosome. Patients with an XXY genotype often have classic Klinefelter syndrome; those with an XXXY, XXXXY, or XXYY genotype or with XXY/chromosomal mosaicism may have milder, variant forms of the syndrome.

Etiology and Pathophysiology

The XXY genotype is usually due to meiotic nondisjunction during parental gametogenesis that results in an egg with two X chromosomes or a sperm with an X and a Y chromosome. After fertilization, nondisjunction during mitotic division results in mosaicism.

Infants and young boys with Klinefelter syndrome may present with micropenis, but there are commonly no physical stigmata of Klinefelter syndrome until puberty. As serum and testicular gonadotropin concentrations increase at the usual age of puberty, the seminiferous tubules do not enlarge, but instead undergo fibrosis and hyalinization that results in small, firm testes. Obliteration of the seminiferous tubules results in azoospermia.

In addition to dysgenesis of the seminiferous tubules, the Leydig cells are functionally abnormal. Basal Leydig cell testosterone production rate is reduced, and Leydig cells fail to stimulate normally with endogenous LH or exogenous hCG.

The clinical manifestations of androgen deficiency vary considerably from patient to patient with Klinefelter syndrome and are related to the degree of diminished testosterone production and sometimes to alterations in the androgen receptor. Some untreated boys in their late teens and adult men with Klinefelter syndrome have virtually no secondary sexual developmental changes, but others are indistinguishable from healthy individuals. Based on data from the Danish national registry, we know that only 25% of men with Klinefelter syndrome are correctly diagnosed. Although many of the missed diagnoses are due to failure to recognize Klinefelter syndrome, the high percentage of missed diagnoses is due to minimally abnormal phenotype of many men with Klinefelter syndrome.

Klinefelter syndrome is characterized by a compensatory rise in serum FSH and LH concentrations. The elevated LH concentrations stimulate aromatization of testosterone and testosterone precursor, resulting in increased production of estradiol and estradiol precursors. The relatively high estradiol:testosterone ratio causes variable degrees of gynecomastia in patients with Klinefelter syndrome. The elevated serum estradiol concentrations stimulate the liver to produce more SHBG, resulting in a disproportionately greater decrease in free or unbound testosterone compared to total testosterone. Some men with Klinefelter syndrome have normal total serum testosterone concentrations, but low free serum testosterone concentrations.

In addition to symptoms and signs of male hypogonadism, boys and men with Klinefelter syndrome have an increased prevalence of cognitive impairment. The pathogenesis of the cognitive abnormality is unknown, but it is likely that testosterone deficiency is not the sole cause.

Testicular Pathology

Most of the seminiferous tubules are fibrotic and hyalinized although Sertoli cells and spermatogonia are occasionally present. The tunica propria lack elastic fibers, and the Leydig cells are arranged in clumps and appear hyperplastic although the total mass is normal.

Clinical Features

A. Symptoms and signs   There are usually no symptoms before puberty except poor school performance in some affected individuals. Intellectual impairment is noted in many patients with Klinefelter syndrome, but the true proportion with subnormal intelligence is not known. Deficits in language and higher intellectual functions, such as concept formation and problem solving, have been noted. Poor social skills also are common.

Puberty may be delayed but rarely more than 1 to 2 years. During puberty, the penis and scrotum undergo varying degrees of development, with some individuals appearing normal. Most patients with classic Klinefelter syndrome (80%) have diminished facial and torso hair growth. The major complaint is persistent gynecomastia in the majority of patients. Teenage boys and men are often embarrassed by incomplete genital development. Other symptoms include decreased libido (compared to peers) and infertility. The testes are classically small (<4 mL each in volume) and very firm. Bone mineral density may be dramatically reduced and osteoporotic fractures are more common in patients with long-standing, untreated androgen deficiency.

Patients with Klinefelter syndrome have abnormal skeletal proportions that are similar to, but not identical, to eunuchoidal proportions. Growth of the legs is relatively greater than that of the trunk and arms; therefore, pubis-to-floor height (which represents appendicular bone growth) is greater than crown-to-pubis height (which predominantly represents axial bone growth), and arm span is less than total height (Figure 12–7). Thus, the abnormal skeletal proportions are not solely the result of androgen deficiency (which results in arm span greater than height).

Several clinical and genotypic variants of Klinefelter syndrome have been described. In addition to small testes with seminiferous tubular hyalinization, azoospermia, deficient secondary sexual development, and elevated gonadotropins, patients with three or more X chromosomes uniformly have severe cognitive impairment.

Skeletal deformities such as radioulnar synostosis, flexion deformities of the elbows, and clinodactyly are more commonly seen in Klinefelter syndrome. Patients with sex chromosome mosaicism (XX/XXY) typically manifest a mild phenotype. These patients may have normal testicular size and may be fertile if their testes contain the XY genotype.

Klinefelter syndrome is associated with a shortened overall life expectancy (by a few years) and a small absolute increased risk of autoimmune disease, diabetes mellitus, chronic pulmonary disease (emphysema, chronic bronchitis), extragonadal germ cell tumors, deep venous thrombosis and other medical diseases and comorbidities. There is a significantly increased relative lifetime risk of breast cancer, but the absolute lifetime risk remains very low.

FIGURE 12–7 Klinefelter syndrome in a 20-year-old man. Note relatively increased lower/upper body segment ratio, gynecomastia, small penis, and sparse body hair with a female pubic hair pattern.

B. Laboratory findings Serum testosterone is typically low or low normal, and serum FSH and LH concentrations are elevated with FSH concentrations always more elevated than LH concentrations. Men with classic Klinefelter syndrome are azoospermic although some men with mosaic Klinefelter syndrome have variable degrees of oligozoospermia. Buccal smear analysis for examination of Barr bodies is not commonly done because this test is not sensitive enough. Serum karyotyping reveals multiple X chromosomes, but even a normal serum karyotype does not definitively exclude Klinefelter syndrome.

Differential Diagnosis

Klinefelter syndrome should be distinguished from other causes of hypogonadism. The finding of small, firm testes and elevated serum gonadotropin concentrations is virtually pathognomonic of Klinefelter syndrome. Other causes of primary hypogonadism with testosterone deficiency include mumps orchitis (particularly postpubertal mumps orchitis), severe pelvic trauma, testicular ischemia and orchidectomy. These other causes of primary hypogonadism are usually obvious by history.

Treatment

A. Medical treatment Teenage boys and men with Klinefelter syndrome with low or low-normal serum testosterone concentrations should be treated with testosterone replacement. There is some controversy about the timing of initiation of testosterone therapy in men with Klinefelter syndrome and normal serum testosterone concentrations. Teenage boys should be treated with low

dosages of testosterone to allow for some degree of virilization but to avoid stunting adult height due to sex steroid-induced premature closure of bony epiphyses. Testosterone enanthate or cypionate, 50 to 100 mg intramuscularly, monthly or low dosage testosterone gel to deliver 2 to 3 mg of testosterone daily is sufficient to virilize teenage boys. The authors prefer monitored administration of intramuscular testosterone enanthate or cypionate in teenage boys because this therapy is less expensive and ensures adherence. The dosage of testosterone replacement should be increased to adult dosages when growth has ceased or by age 18.

B. Surgical treatment If gynecomastia presents a significant cosmetic problem, reduction mammoplasty may be performed. Mammoplasty should be delayed until after androgen therapy has been optimized (in order to prevent recurrent gynecomastia after surgery). A surgeon who is experienced at the procedure should be consulted in order to optimize cosmetic results.

CRYPTORCHIDISM

Cryptorchidism is unilateral or bilateral absence of the testes from the scrotum because of failure of normal testicular descent from the genital ridge through the external inguinal ring. Between 1% and 8% of full-term and 20% to 25% of premature male infants have cryptorchidism. In most cases of cryptorchidism noted at birth, spontaneous testicular descent occurs during the first year of life, reducing the incidence to 0.2% to 0.8% at 1 year of age. Unilateral cryptorchidism is five to ten times more common than bilateral cryptorchidism.

Almost 50% of cryptorchid testes are located at the external inguinal ring or in a high scrotal position; about 20% lie within the inguinal canal between the internal and external inguinal rings; about 10% are intra-abdominal; and about 25% are ectopic (ie, located away from the normal pathway of descent from the abdominal cavity to the scrotum). Most ectopic testes are found in a superficial inguinal pouch above the external inguinal ring.

Etiology and Pathophysiology

Testicular descent usually occurs between the 12th week of fetal development and birth. Both mechanical and hormonal factors appear to be important for this process. Cryptorchidism is more common in patients with congenital defects in androgen synthesis or action and in patients with congenital gonadotropin deficiency, and experimental studies have demonstrated that DHT is required for normal testicular descent. These observations suggest that prenatal androgen deficiency contributes to the development of cryptorchidism. Environmental endocrine disruptors with estrogen or antiandrogen activity, such as phthalates or bisphenol A, have been implicated as causes of cryptorchidism.

Pathology

Histological studies on cryptorchid testes have demonstrated a decrease in the size of the seminiferous tubules and number of spermatogonia and an increase in peritubular tissue. The Leydig

cells usually appear normal. Abnormalities have been detected as early as 6 months. It is well established that the longer a testis remains ectopic, the more likely it is to show pathologic changes. More severe changes are generally found in intra-abdominal testes than inguinal testes.

It is not known whether pathologic changes in the testes are due to the effects of cryptorchidism or to intrinsic abnormalities in the gonad. Serial testicular biopsies in cryptorchid patients have demonstrated partial reversal of the histologic abnormalities following orchiopexy, suggesting that the environment is partly responsible for the observed pathologic abnormalities. Scrotal testes are significantly cooler (1.5-2°C) than intra-abdominal temperatures; one hypothesis is that higher temperatures contribute to testicular damage.

An intrinsic abnormality in the testes in patients with unilateral cryptorchidism is suggested by the observation that such patients are at increased risk for development of germ cell neoplasms in the scrotal (eutopic) testis.

Clinical Features

A. Symptoms and signs   There are usually no symptoms unless a complication such as testicular torsion, trauma, or malignant degeneration occurs.

Absence of one or both testes is the cardinal clinical finding. This may be associated with a small scrotum (bilateral cryptorchidism) or hemiscrotum (unilateral cryptorchidism). Signs of androgen deficiency are not present.

B. Laboratory findings   Serum AMH, testosterone, FSH, and LH concentrations are normal for age in prepubertal boys with bilateral or unilateral cryptorchidism. Although the diagnosis should be made in early childhood, postpubertal boys, and men with unilateral or bilateral cryptorchidism have normal serum testosterone and LH concentrations, but may have an isolated elevation of serum FSH due to damage to seminiferous tubules and decreased Sertoli cell secretion of inhibin B.

C. Imaging studies   When examined by an experienced clinician, at least 70% of cryptorchid testes are palpable in the inguinal canal. Therefore, imaging is not usually necessary. Ultrasound is not generally useful. Gadolinium-infusion magnetic resonance venography is a sensitive method for localizing the nonpalpable cryptorchid testes, but it is expensive and may require sedation for young children.

Differential Diagnosis

Cryptorchidism must be discriminated from retractile testis (pseudocryptorchidism) that is due to a hyperactive cremasteric reflex. Cold temperature, fear, and genital manipulation commonly activate the reflex, which is most prominent around the age of 5 years. The child should be examined with warm hands in a warm room. Application of a warm towel on the scrotal sac may also allow the testis to descend spontaneously. The testis can usually be milked into the scrotum with gentle pressure over the lower abdomen in the direction of the inguinal canal.

The virilizing forms of congenital adrenal hyperplasia may result in prenatal fusion of the labial-scrotal folds and clitoral hypertrophy (Chapter 14). Females with virilizing congenital adrenal hyperplasia have the appearance of phenotypic males with micropenis and bilateral cryptorchidism. Because of the potentially disastrous consequences (acute adrenal insufficiency) if this diagnosis is missed, male infants with micropenis and bilateral cryptorchidism should always be evaluated for a virilizing form of congenital adrenal hyperplasia.

Congenital bilateral anorchia (discussed later) is associated with low serum AMH concentrations and an absent or subnormal response to stimulation with hCG.

Complications and Sequelae

Cryptorchidism increases the risk of inguinal hernia (due to failure of the processus vaginalis to close) and increases the risk of torsion, but the two most important complications of untreated cryptorchidism are testicular malignancy and infertility.

A. Testicular germ cell malignancy   Cryptorchidism increases the risk of testicular cancer by 3 to 8 fold compared to the risk in boys with normally descended testes. The incidence of such tumors is greater in patients with intra-abdominal testes than in patients with inguinal testes. Seminomas are the neoplasms most commonly associated with maldescended testes. In addition, there is an increased risk for development of testicular intraepithelial neoplasia (carcinoma in situ). Surgical correction with orchiopexy before puberty appears to modestly reduce the risk of malignancy.

Because of the increased risk of neoplasia, many urologists recommend orchidectomy for a unilaterally undescended testicle in a patient first seen after puberty. Patients who present with bilateral cryptorchidism after puberty should have bilateral orchiopexy and testicular biopsies to assess for neoplasia, preserve testicular endocrine function and make palpation for detection of neoplasia easier.

B. Infertility   Over 75% of untreated bilaterally cryptorchid men are infertile. Spermatogenesis and fertility are often reduced in men with untreated unilateral cryptorchidism, but men with a history of unilateral cryptorchidism that was surgically treated early in life generally have normal fertility.

Treatment

Most testes that are undescended at birth enter the scrotum during the first year of life, and the majority descend during the first 3 months of life. It is rare for a cryptorchid testis to descend spontaneously after the age of 6 months. Because germ cell damage begins very early with rapid decreases in germ cell number after age 12 months, surgical correction of an undescended testis should be performed between ages 6 and 12 months.

A. Medical   Medical treatment with hCG or GnRH agonists was formerly recommended. However, the long-term success rates

of medical therapy are very low compared to orchiopexy. Randomized, blinded studies of hCG demonstrate success rates of ~5% to 20%. GnRH agonists might be marginally more effective, but the best studies demonstrate success rates of only ~5% to 40%. Because the long-term success rate is low and because medical therapy delays definitive surgical therapy with a risk for further germ cell loss, medical therapy is generally not recommended.

B. Surgical treatment   Several procedures have been devised to place the maldescended testis into the scrotum (orchiopexy). The operation may be performed in one or two stages. Inguinal hernia should be repaired if present. The surgery should be performed between 6 and 12 months of age.

CONGENITAL BILATERAL ANORCHIA (VANISHING TESTES SYNDROME)

Congenital bilateral anorchia is found in approximately 1 of 20,000 boys.

Etiology and Pathophysiology

Normal functional testicular tissue (producing normal amounts of fetal testosterone) must be present during the first 14 to 16 weeks of embryonic development in order for Wolffian duct growth and Müllerian duct regression to occur and for development of male external genitalia. Absence of normal fetal testosterone concentration in a male fetus during the first trimester results in varying degrees of ambiguous genitalia. Prenatal testicular injury occurring after 16 weeks of gestation as a result of trauma, vascular insufficiency, infection, or other mechanisms may result in loss of testicular tissue ("vanishing testes") in an otherwise normal phenotypic male.

Testicular Pathology

Wolffian duct structures are generally normal, and the vas deferens and testicular vessels may terminate blindly or in a mass of connective tissue in the inguinal canal or scrotum.

Clinical Features

A. Symptoms and signs   At birth, patients appear to be normal phenotypic males but with an empty scrotal sac and no palpable inguinal testes. Growth and development are normal until secondary sexual development fails to occur at puberty. The penis remains small; pubic and axillary hair does not fully develop despite the presence of adrenal androgens; and the scrotum remains empty. If a teenage patient does not receive androgens, eunuchoidal proportions develop.

There are a few case reports of boys undergoing partial spontaneous virilization at puberty. Although anatomically no testicular tissue has been identified in these boys, venous catheterization studies have demonstrated higher testosterone concentrations in venous blood obtained from the spermatic veins than in the peripheral venous circulation. This suggests that functional Leydig cells are present in these rare cases, but these boys have still lacked testicular germinal epithelium or stroma (and therefore cannot produce sperm).

B. Laboratory findings   In boys under age 12 months, teenage boys and men, serum testosterone and AMH concentrations are very low, and both LH and FSH are markedly elevated. Serum testosterone concentrations do not rise following an hCG stimulation test. Chromosomal analysis discloses a 46,XY karyotype.

Differential Diagnosis

Although congenital bilateral anorchia is very rare, it must be distinguished from bilateral cryptorchidism, a much more common disorder. The most commonly used test is measurement of serum AMH. Serum AMH is low in bilateral anorchia and normal in cryptorchidism. The hCG stimulation test is sometimes useful and will demonstrate a normal response in cryptorchidism, but will demonstrate no rise in serum testosterone concentration in congenital bilateral anorchia.

Treatment

A. Medical treatment   Medical treatment is similar to that outlined for patients with Klinefelter syndrome.

B. Surgical treatment   Implantation of testicular prostheses for cosmetic purposes may be beneficial after the scrotum has enlarged in response to androgen therapy.

LEYDIG CELL APLASIA

Defective development of testicular Leydig cells is a very rare cause of male pseudohermaphroditism with ambiguous genitalia.

Etiology and Pathophysiology

Testes are present in the inguinal canals and contain prepubertal-appearing tubules with Sertoli cells and spermatogonia without germinal cell maturation, and there are no or few Leydig cells. The syndrome is caused by inactivating mutations in the LH receptor that alters receptor signal transduction. The presence of vasa deferens and epididymi and absence of Müllerian structures (eg, uterus) in these patients indicates that the local concentration of testosterone and AMH was high enough during embryogenesis to result in differentiation of the Wolffian duct structures and prevent development of Müllerian structures. However, the ambiguity of the genitalia indicates that the androgen concentration in these patients was insufficient to bring about full virilization of the external genitalia.

Clinical Features

A. Symptoms and signs   These patients may present in infancy with variable degrees of genital ambiguity, including a

bifid scrotum, clitoral phallus, urogenital sinus, and blind vaginal pouch. Alternatively, they may appear as normal phenotypic females and escape detection until adolescence, when they present with primary amenorrhea, with or without normal breast development. The gonads are generally located in the inguinal canal. Axillary and pubic hair, although present, may be sparse. Milder defects in the LH receptor may result in Leydig cell hypoplasia with micropenis, hypospadias, and reduced fertility.

B. Laboratory findings   Serum testosterone levels are low (in the female range) and serum gonadotropins are elevated. There is no increase in testosterone following hCG administration.

Differential Diagnosis

Patients with mild forms of Leydig cell aplasia must be distinguished from cryptorchidism and vanishing testes syndrome. Low serum AMH concentrations and lack of response to hCG stimulation test distinguishes Leydig cell aplasia from cryptorchidism. Finding testicular tissue by palpation, imaging or surgical exploration discriminates Leydig cell aplasia from vanishing testes syndrome. More severe forms of Leydig cell aplasia that present with ambiguous genitalia must be discriminated from testosterone biosynthetic defects, disorders of androgen action, and 5α-reductase deficiency; these disorders are discussed in Chapter 14.

Treatment

Patients with mild Leydig cell aplasia who have been raised as boys may respond well to testosterone replacement therapy as outlined for Klinefelter syndrome although genital development and spermatogenesis do not normalize. The management of patients with more severe Leydig cell aplasia and ambiguous genitalia is more controversial; see Chapter 14.

NOONAN SYNDROME (MALE TURNER SYNDROME)

Phenotypic and genotypic males with many of the physical stigmas of classic Turner syndrome have been described under a variety of names, including Noonan syndrome and male Turner syndrome. Noonan syndrome has an incidence of 1:1000 to 1:2500 live births and may occur sporadically or in families. It is inherited in an autosomal dominant fashion with variable penetrance. It is a RASopathy, a syndrome due to a germline mutation affecting the RAS/mitogen-activated protein kinase (MAPK) pathway. Noonan syndrome is associated with a number of structural and functional abnormalities of the male reproductive tract: cryptorchidism, delayed puberty, and decreased spermatogenesis as adults.

Clinical Features

A. Symptoms and signs   The most common clinical features are short stature, webbed neck, hypertelorism, and lentigines. Congenital cardiac anomalies are common and involve primarily the right side of the heart—in contrast to patients with XO gonadal dysgenesis (Turner syndrome).

Cryptorchidism is frequently present. Testicular size is generally small during puberty but gradually grows to normal volumes by adulthood.

B. Laboratory findings   Serum testosterone, LH, and FSH concentrations are usually normal before and during puberty. Serum testosterone concentrations are usually normal, but serum LH and FSH may be slightly elevated in adult men with Noonan syndrome. The karyotype is 46,XY.

Differential Diagnosis

The clinical features of Noonan syndrome are usually sufficiently distinctive to preclude other diagnoses.

Treatment

Treatment of cryptorchidism and hypogonadism is as outlined in the sections under cryptorchidism and Klinefelter syndrome.

CAUSES OF PRIMARY HYPOGONADISM PRESENTING IN ADULTHOOD

MYOTONIC DYSTROPHY

Myotonic dystrophies types 1 and 2 are autosomal dominant disorders of muscle with multisystem clinical features including primary hypogonadism.

Testicular histology varies from moderate derangement of spermatogenesis with germinal cell arrest to regional hyalinization and fibrosis of the seminiferous tubules. The Leydig cells are usually preserved and may appear in clumps.

The testes are normal in affected prepubertal individuals, and puberty generally proceeds normally. After puberty, seminiferous tubular atrophy results in a decrease in testicular size and change of consistency from firm to soft or mushy. Infertility is a consequence of disrupted spermatogenesis. If testicular hyalinization and fibrosis are extensive, Leydig cell function may also be impaired.

Clinical Features

A. Symptoms and signs   The disease usually becomes apparent in adulthood. Progressive weakness and atrophy of the facial, neck, hand, and lower extremity muscles is commonly observed. Severe atrophy of the temporalis muscles, ptosis due to weakness of the levator muscles of the eye with compensatory wrinkling of the forehead muscles, and frontal baldness comprise the characteristic myopathic facies. Myotonia is present in several muscle groups and is characterized by inability to relax the muscle normally after a strong contraction.

Testicular atrophy is not noted until adulthood, and most patients develop and maintain normal facial and body hair growth and libido. Gynecomastia is usually not present.

Associated features include mental retardation (type 1 disease only), cataracts, diabetes mellitus, cardiac arrhythmias, and primary hypothyroidism.

B. Laboratory findings Serum testosterone is normal to slightly decreased. FSH is elevated in the majority of patients. LH is also frequently elevated, even in patients with normal serum testosterone levels. Leydig cell reserve is generally diminished, with subnormal increases in serum testosterone following stimulation with hCG.

Treatment

There is no therapy that will prevent progressive muscular atrophy in this disorder. Testosterone replacement therapy is not indicated unless serum testosterone levels are subnormal.

LATE-ONSET MALE HYPOGONADISM

In contrast to the menopause in women, men do not experience a sharp decline in gonadal function. However, a gradual diminution of testosterone production occurs in many men as part of the aging process. It is not known how many men develop symptoms directly attributable to this phenomenon. The decline in serum testosterone concentrations in older men without a specific identifiable cause other than aging has been described as late-onset hypogonadism. Andropause and male climacteric are inaccurate and misleading descriptors because these terms imply that testicular function declines rapidly in aging men. Even the term "late-onset hypogonadism" is problematic because the term implies that such men are hypogonadal. Low serum testosterone concentration is likely a marker of poor health and not an indicator of testosterone deficiency in many older men.

Etiology, Pathology, and Pathophysiology

During middle age, serum total and free testosterone concentrations begin to decline. The levels of free testosterone decrease to a greater extent because of an increase in SHBG. Leydig cell testosterone production rates decline as Leydig cell responsiveness to LH decreases. Aging also is associated with perturbations in hypothalamo-pituitary function that result in lower LH secretion than normal for a given serum testosterone concentration. As a result of the dual defect in testicular and hypothalamo-pituitary function, aging men tend to have lower serum testosterone concentrations and higher LH concentrations than younger men, but serum testosterone and LH concentrations tend to remain in the low-normal and high-normal range, respectively. Older men often lose the normal circadian rhythm of testosterone secretion; in older men, early serum morning testosterone concentrations might not be significantly higher than late afternoon or early evening serum testosterone concentrations. On the other hand, healthy older men may maintain a normal circadian rhythm and

normal serum testosterone concentrations. Thus, some experts contend that declining hypothalamo-pituitary-testicular function is a marker of declining overall health and not an inevitable physiological effect of aging.

Histologic studies of the aging testes have shown patchy degenerative changes in the seminiferous tubules with a reduction in number and volume of Leydig cells. The pathologic changes are first noted in the regions most remote from the arterial blood supply. In some men, microvascular insufficiency may be the etiologic basis for the histologic tubular changes and the decrease in Leydig cell function noted with aging. In addition, virtually all of the conditions (eg, infection, tumor, infiltrative disease, and vasculopathy) that cause adult seminiferous tubule failure may lead to Leydig cell dysfunction if testicular injury is severe enough.

Clinical Features

A. Symptoms and signs Many symptoms have been attributed to late-onset hypogonadism, including decreased libido, erectile dysfunction, depressed mood, fatigue, decreased strength, and decreased cognitive function and mental concentration. The most specific symptoms of late-onset hypogonadism are sexual dysfunction: decline in libido, sexual fantasies, and early morning erections.

B. Laboratory findings In older men, serum testosterone concentration is often slightly low or low-normal. Serum LH concentration is usually high-normal or slightly above the normal limit, but a number of older men have a borderline low total and free testosterone concentration and a LH concentration that is well within the normal range. Sperm concentrations are often lower in older men. Bone mineral density may be low, but it is unclear whether this finding is causally related to declining serum testosterone production.

Differential Diagnosis

In older men with low serum testosterone concentrations who present with nonspecific symptoms such as decreased energy, depressed mood, sexual dysfunction or weakness, causes other than male hypogonadism must be considered. Depression and systemic illness are often the cause of these symptoms. In men with low serum testosterone concentrations and osteoporosis, other causes of osteoporosis must be excluded before attributing hypogonadism as the cause.

Treatment

There is great controversy about whether and how to treat older men with late onset hypogonadism. There are very little data from randomized, placebo-controlled studies to determine the benefit and risk of testosterone in such men. The best data to date come from early findings from the Testosterone Trial, a large study of men, aged greater than or equal to 65 years, with slightly low serum testosterone concentrations (average of two early morning serum total testosterone concentrations ≤275 ng/dL) and

symptoms of male hypogonadism who were randomized to 12 months of testosterone gel, at a dosage sufficient to raise testosterone concentrations into the normal range, or placebo gel. This study demonstrated a small, but significant benefit in sexual function and physical function in men randomized to testosterone gel. Although there is controversy about the clinical significance of these improvements, testosterone therapy improved multiple aspects of sexual function (including sexual desire, erections, and orgasm frequency). Men who were randomized to testosterone gel reported having an average of about four orgasms per month compared to an average of about three orgasms per month in men in the placebo gel group. In the physical function assessment, almost twice as many men in the testosterone gel group (20.5% vs 12.6%) increased their 6-minute walk test by a clinically important amount (≥50 meters over baseline in 6 minutes; average baseline walk test = approximately 390 meters). Men treated with testosterone also reported improved mood and energy compared to men treated with placebo.

The major safety concerns for testosterone therapy are potential adverse effects on cardiovascular events and prostate cancer. In the 12-month Testosterone Trial, there was no significant difference in adverse cardiovascular events or prostate cancer. There were seven deaths in the placebo group and three in the testosterone group. Men randomized to testosterone gel were more likely to develop erythrocytosis (seven in the testosterone group vs zero in the placebo group) in this study.

At this time, there is insufficient evidence to make a general recommendation about the management of putative late-onset hypogonadism. The authors recommend the following: before considering testosterone therapy for older men with low serum testosterone concentrations and symptoms of hypogonadism, the clinician must (1) confirm that serum testosterone concentrations are consistently low (in two or more early morning blood samples using a reliable testosterone assay); (2) optimize treatment of depression and systemic disease and reduce or eliminate medications that might cause symptoms such as fatigue, sexual dysfunction or altered mood (eg, corticosteroids, opioids, and medications that affect the central nervous system); and (3) assess for common causes of secondary hypogonadism (including Cushing syndrome and hyperprolactinemia; see section on Evaluation of Male Hypogonadism). If the man continues to have symptoms that suggest hypogonadism, serum testosterone concentrations continue to be below the normal range, and no specific cause of secondary hypogonadism has been identified, then the clinician should discuss the potential benefits (improved sense of well-being, improved sexual function, and improved strength) and risks (erythrocytosis and the possibility, albeit unproven and controversial, of increased risk of cardiovascular events and prostate cancer). Before initiating a trial of testosterone therapy for "late onset of hypogonadism," the clinician should document this discussion in the medical record. Some experts recommend that older men sign an "informed consent form" documenting a discussion of the potential benefits and risks of testosterone therapy. If the patient has no improvement of symptoms or objective evidence of improvement (eg, improved bone mineral density), then the clinician should consider discontinuation of testosterone therapy. A 6-month trial of testosterone is sufficient for improvement of symptoms, but a longer trial (up to 2 years) might be necessary to assess improvement in bone mineral density.

SPECIFIC SEQUELAE OF HYPOGONADISM

MALE INFERTILITY

About 15% of married couples have difficulty in conceiving. In about 30% of such couples, only the man has reproductive dysfunction, and an additional 20%, both partners have a reproductive abnormality. In the United States, infertility is classically defined as the inability to conceive after 1 year of frequent vaginal intercourse without the use of contraception. About 50% of young, healthy couples that fail to conceive after 1 year of unprotected intercourse will conceive in the subsequent year. Subfertility is a term used to describe reduced fertility that might require assisted reproductive technology (ART) such as surgical extraction of sperm and ova followed by ICSI into an ovum. Subfertility also describes couples who take longer than 12 months to conceive, but would conceive eventually without treatment. ART is an expensive method to hasten conception in such couples.

Etiology and Pathophysiology

In order for conception to occur, spermatogenesis must be normal, the sperm must complete its maturation and transportation through patent ducts, adequate amounts of seminal plasma must be added to provide volume and nutritional elements, and the male must be able to deposit the semen near the female's cervix. Any defect in this pathway can result in subfertility due to a male factor problem. Spermatozoa must also be able to penetrate the cervical mucus and reach the uterine tubes, where conception takes place.

Table 12–5 lists the identified causes of male infertility. Most male infertility and subfertility are due to primary testicular dysfunction. The most common cause of male subfertility (35%-45%) is due to idiopathic primary testicular dysfunction with abnormal spermatogenesis (idiopathic spermatogenic failure). Idiopathic spermatogenic failure is associated with sperm concentrations less than 10 million/mL (normal >15 million/mL) and is usually associated with abnormal sperm morphology and motility. Men with idiopathic spermatogenic failure have normal serum testosterone and LH concentrations, but many have an isolated elevation in serum FSH.

About 10% to 15% of men with infertility and absent (azoospermia) or very low sperm concentrations (<1 million/mL) have Klinefelter syndrome with low or low-normal serum testosterone and elevated serum gonadotropin concentrations. Other identifiable causes of male infertility and subfertility due to primary testicular dysfunction include genetic mutations (eg, Y chromosome microdeletions), drugs (particularly systemic chemotherapy for neoplastic disease), testicular cancer, cryptorchidism, infection

TABLE 12-5 Causes of male infertility.

Endocrine
　Hypothalamic-pituitary disorders
　Primary testicular disorders
　Defects of androgen action
　Hyperthyroidism
　Hypothyroidism
　Adrenal insufficiency
　Congenital adrenal hyperplasia
　Systemic illness (causing hypogonadotropism)
Defects in spermatogenesis
　Idiopathic spermatogenic failure
　Y chromosome microdeletions
　Drug induced (eg, alkylating chemotherapy)
　Immotile cilia syndrome
Sperm transportation defects
　Ductal obstruction
　　Congenital (eg, cystic fibrosis)
　　Acquired (eg, due to infection)
　Retrograde ejaculation
Sexual function disorders
　Infrequent vaginal intercourse
　Erectile dysfunction
　Anatomic defects of the penis
Miscellaneous
　Antibodies to sperm or seminal plasma
　Varicocele

(eg, postpubertal mumps orchitis), and pelvic irradiation or surgery. Abuse of tobacco, marijuana, or alcohol is also associated with gonadotoxicity, defective spermatogenesis and subfertility. There is significant controversy about whether varicoceles affect male fertility. Varicoceles are commonly found in subfertile men (25%-40%), but they are also commonly found in men with normal fertility (8%-20%). Large varicoceles might impair spermatogenesis by a variety of mechanisms including compromised testicular cooling and increased intratesticular reactive oxygen species. Autoimmune disturbances that lead to sperm agglutination and immobilization are very rare.

Causes other than primary testicular defects in spermatogenesis include sexual disorders, endocrinopathies that affect the hypothalamo-pituitary axis, and defects in sperm transportation.

Sexual disorders include erectile dysfunction and failure to have frequent vaginal intercourse. These are uncommon causes of infertility and are relatively easy to remedy. To optimize the likelihood of conception, the frequency of vaginal intercourse should be at least two to three times weekly. A very small percentage (<5%) of male subfertility is due to disturbances in hypothalamic, pituitary, adrenal, or thyroid function. Defects in sperm transportation include obstruction of the ejaculatory tract, retrograde ejaculation, or anejaculation. An important cause of ejaculatory tract obstruction is congenital bilateral absence of the vasa deferentia. The majority of men with congenital bilateral absence of the vasa deferentia have mutations of the cystic fibrosis gene. Nearly all men with clinical cystic fibrosis have congenital bilateral absence of the vasa deferentia, but congenital bilateral absence of the vasa deferentia may be the only clinical manifestation of a cystic fibrosis gene mutation.

Clinical Features

A. Symptoms and signs   Men with idiopathic spermatogenic failure (and preserved Leydig cell function) will have no symptoms other than infertility, but men with combined spermatogenic failure and Leydig cell dysfunction may have symptoms of hypogonadism. Most men with infertility due to endocrinopathy will have symptoms related to the endocrinopathy; the clinician should query about symptoms of thyroid dysfunction, corticosteroid excess, or symptoms of a large sellar mass (headaches and visual changes).

B. Laboratory findings   The single most useful test is seminal fluid analysis. Because of significant variation in sperm concentrations, collection an analysis of at least two seminal fluid samples is essential. The sample should be collected after 2 to 7 days of ejaculatory abstinence and analyzed by a laboratory that follows the World Health Organization standards. A normal report indicates normal endocrine function and spermatogenesis and an intact transport system.

Measurement of an early morning serum testosterone, FSH and LH concentrations should be performed, and abnormal results confirmed with repeat testing. Patients with severe oligozoospermia or azoospermia and elevated serum LH and FSH concentrations may have Klinefelter syndrome and should be offered testing for a karyotypic abnormality. In addition, some experts advise testing for Y chromosome deletions in all men with oligozoospermia or azoospermia before treatment with assisted reproductive technology. If vasa deferentia are not palpable or visible on ultrasound, then testing for cystic fibrosis transmembrane conductance regulator gene mutations should be performed.

The female partner should be thoroughly examined to verify patency of the uterus and oviducts, normal ovulation, and normal cervical mucus. This examination must be performed even in the presence of a male factor abnormality, because infertility is due to a combination of male and female factors in about 20% of cases.

C. Imaging studies   In men with low seminal fluid volumes, azoospermia and normal serum testosterone, LH and FSH concentrations, a transrectal ultrasound should be performed to look for ejaculatory duct obstruction. Testicular ultrasound should be performed in men with scrotal masses if the physical examination of the scrotum is difficult or incomplete or if a large varicocele is suspected.

Treatment

A. Endocrine disorders   Correction of hyperthyroidism, hypothyroidism, adrenal insufficiency, and congenital adrenal hyperplasia generally restores fertility. Gonadotropin therapy is often effective in stimulating spermatogenesis in patients with hypogonadotropic hypogonadism. Gonadotropin therapy is initiated with LH replacement therapy in the form of hCG (1000-2000 units subcutaneously two to three times weekly) for 6 months. FSH replacement (75-150 IU units subcutaneously three times per week) is added after 6 months if sperm cells do not

appear in the ejaculate. Gonadotropin therapy often does not normalize sperm concentrations, but may still allow impregnation. Men with postpubertal onset of gonadotropin deficiency may respond to hCG monotherapy.

B. Defects of spermatogenesis   Drugs that interfere with spermatogenesis should be discontinued. These include marijuana, alcohol, monoamine oxidase inhibitors, sulfasalazine, and nitrofurantoin. Many men with oligozoospermia and even azoospermia may conceive with use of ART including *in vitro* fertilization, surgical sperm extraction, and ICSI.

C. Ductal obstruction   Localized obstruction of the vas deferens may be treated by vasovasostomy. Sperm are detected in the ejaculate of 60% to 80% of patients following this procedure. However, the subsequent fertility rate is only 40% to 50%; the presence of anti-sperm antibodies that agglutinate or immobilize sperms probably accounts for the high failure rate.

Epididymovasostomy may be performed for epididymal obstruction. Sperm in the postoperative ejaculate have been found in approximately half of patients treated with this procedure, but subsequent fertility has been demonstrated in only 20% of cases. ART is often used in cases of ductal obstruction.

D. Varicocele   Although there are few data to support treatment of varicoceles to improve fertility, surgical ligation, or radiographic embolization of the incompetent spermatic veins is sometimes performed in men with infertility and large varicoceles.

E. Retrograde ejaculation   Retrograde ejaculation of semen into the urinary bladder may occur following disruption of the internal bladder sphincter or with neuropathic disorders such as diabetic autonomic neuropathy. Phenypropanolamine may improve antegrade ejaculation in patients with retrograde ejaculation. Sperm can also be recovered from urine following masturbation and used for direct intrauterine insemination or other ART procedures.

Course and Prognosis

For male subfertility due to hypogonadotropic hypogonadism and other endocrinopathies, the prognosis for fertility is good. For all other causes of male factor infertility, ART, though expensive, makes it possible for many couples to conceive. For many couples, adoption might be the best option, but adoption may also be expensive and time consuming.

ERECTILE DYSFUNCTION

Erectile dysfunction is the inability to achieve or maintain an erection of sufficient duration and firmness to complete satisfactory sexual activity in more than 25% of attempts. It may occur with or without associated disturbances of libido or ejaculation. Approximately 5% of men have complete erectile dysfunction by age 40 and 15% by age 70. Some degree of erectile dysfunction is present in about 50% of men between ages 40 and 70.

Etiology and Pathophysiology

Penile erection occurs when blood flow to the penile erectile tissue (corpora cavernosa and spongiosum) increases as a result of dilation of the urethral artery, the artery of the bulb of the penis, the deep artery of the penis, and the dorsal artery of the penis following psychogenic or sensory stimuli transmitted to the limbic system and then to the thoracolumbar and sacral autonomic nervous system. The relaxation of the cavernosal arterial and cavernosal trabecular sinusoidal smooth muscle occurs following stimulation of the sacral parasympathetic (S2-4) nerves, which results in the release of acetylcholine, vasoactive intestinal peptide, and an endothelial cell-derived nitric oxide, which activates guanylyl cyclase. Activation of guanyl cyclase and production of intravascular cycle guanosine monophate results in penile arterial vasodilation and significant increases in blood flow to the penis. As penile blood flow increases, the penile sinusoids become engorged and cause compression of the subtunical venous plexus against the tunica albuginea. The compression of the subtunical venous plexus prevents egress of blood from the penis, resulting in very high penile blood pressures and tumescence. Contraction of the bulbocavernosus muscle through stimulation of the somatic portion of the S2 to S4 pudendal nerves further increases the intracavernosal pressure and tumescence. These processes result in the distention, engorgement, and rigidity of the penis that constitute erection.

Broadly speaking, erectile dysfunction may be divided into psychogenic and organic causes. Major epidemiologic factors that have been associated with erectile dysfunction include diabetes mellitus, metabolic syndrome, hypertension, smoking, aging, and cardiovascular disease. Table 12–6 lists various pathologic conditions and drugs that may be associated with erectile dysfunction.

Most organic causes of erectile dysfunction result from disturbances in the neurologic pathways essential for the initiation and maintenance of erection or in the blood supply to the penis. Many of the endocrine disorders, systemic illnesses, and drugs associated with erectile dysfunction affect libido, the autonomic pathways essential for erection, or the blood flow to the penis. Venous incompetence because of anatomic defects in the corpora cavernosa or subtunical venous plexus is an uncommon cause of erectile dysfunction. Local urogenital disorders such as Peyronie disease (idiopathic fibrosis of the covering sheath of the corpus cavernosum) may mechanically interfere with erection. In some patients, the cause of erectile dysfunction is multifactorial. For example, some degree of erectile dysfunction is reported by over 50% of men with diabetes mellitus who have arteriovascular disease, autonomic neuropathy and other contributors to erectile dysfunction such as depression or medications associated with erectile dysfunction. Male hypogonadism may contribute to worsening erectile function, but it is rarely the sole or primary cause of erectile dysfunction.

Clinical Features

A. Symptoms and signs   Patients may complain of constant or episodic inability to initiate or maintain an erection, and/or decreased penile turgidity. The diagnosis is generally made based on the patient's history. The degree of erectile dysfunction can be

TABLE 12–6 Organic causes of erectile dysfunction.

Neurologic
Brain lesions
Spinal cord lesions
Autonomic neuropathy

Vascular
Pelvic atherosclerosis
Venous leaks

Endocrine
Diabetes mellitus
Hypogonadism
Hyperprolactinemia
Adrenal insufficiency
Feminizing tumors
Hypothyroidism
Hyperthyroidism

Urogenital
Trauma
Irradiation
Peyronie disease

Systemic Illness
Heart failure
Cirrhosis
Uremia
Pulmonary disease with chronic hypoxemia

Postoperative
Aortoilliac or aortofemoral reconstruction
Lumbar sympathectomy
Perineal prostatectomy
Retroperitoneal dissection

Drugs
Endocrinologic
　Anti-androgens
　Estrogens
　5α-Reductase inhibitors
　GnRH analogs
Antihypertensives
　Beta-blockers
Diuretics
　Thiazides
Antidepressants
　Selective serotonin reuptake inhibitors
　Monamine oxidase inhibitors
　Tricyclic antidepressants
Other
　Tobacco
　Alcohol
　Opioids
　Amphetamines
　Cocaine

assessed by questionnaires such as the International Index of Erectile Function or the Sexual Health Inventory for men, but these questionnaires are generally used only in clinical trials. If an underlying neurologic, vascular, or systemic disorder is the cause of erectile dysfunction, additional symptoms and signs referable to the anatomic or metabolic disturbances may be present. A history of claudication of the buttocks or lower extremities indicates atherosclerosis as a likely cause. A thorough history of medication, herb, illicit drug, and alcohol use should be obtained.

The differentiation between psychogenic and organic erectile dysfunction can usually be made on the basis of the history. Even

though the patient may be selectively unable to obtain or maintain a satisfactory erection to complete sexual intercourse, a history of repeated normal erections at other times is indicative of psychogenic erectile dysfunction. Thus, a history of erections that occur nocturnally, during masturbation, or during foreplay or with other sexual partners eliminates significant neurologic, vascular, or endocrine causes of erectile dysfunction. Patients with psychogenic erectile dysfunction often note a sudden onset of sexual dysfunction concurrently with a significant event in their lives such as loss of a friend or relative, an extramarital affair, or the loss of a job.

Patients with organic erectile dysfunction generally note a more gradual and global loss of erectile function. Initially, such individuals may be able to achieve erections with strong sexual stimuli, but ultimately they may be unable to achieve a fully turgid erection under any circumstances. Patients with organic erectile dysfunction may have a concurrent diminution of libido and erectile function; diminished libido may be an indicator of male hypogonadism.

During the physical examination, the patient's secondary sexual characteristics should be assessed and examination performed for gynecomastia, discordant or diminished radial, femoral, or pedal pulses, reduced testicular volume or consistency, penile plaques, and evidence of peripheral or autonomic neuropathy. The bulbocavernosus reflex tests the integrity of the S2 to S4 nerves. It is performed by inserting a finger into the patient's rectum while squeezing his glans penis. Contraction of the anal musculature represents a normal response. Elicitation of the anal wink reflex also tests S2 to S4 nerve function. Visible contraction of anal musculature after stroking the skin around the anus represents a normal response.

B. Laboratory findings and special examinations The diagnosis of erectile dysfunction is made based on the patient's history. Testing of nocturnal penile tumescence is used only in research studies. Most guidelines indicate that laboratory testing for evaluation of erectile dysfunction should be done based on the clinical setting. Although male hypogonadism is seldom the sole or primary cause of erectile dysfunction, it is reasonable to measure serum testosterone concentrations, particularly in any man with low libido and erectile dysfunction. Many experts recommend a fasting lipid profile and serum glucose (or an A1c). These tests are recommended because erectile dysfunction is highly associated with the risk of future cardiovascular events. These tests are not useful for the diagnosis or management of erectile dysfunction; neither improved glycemic control nor treatment of dyslipidemia are effective treatments for erectile dysfunction. The most important assessment is to determine the risk of exertional cardiac ischemia. Typical sexual intercourse requires about three to four metabolic equivalents (METs). This is equal to walking a mile in 20 minutes or walking a flight of 20 steps in 20 seconds or walking up a flight of stairs (at least 10 risers) without pausing or stopping. If the patient with erectile dysfunction cannot perform these activities, then formal cardiac exercise tolerance testing should be considered before any medical therapy for erectile dysfunction is initiated.

The vascular integrity of the penis may be examined by Doppler ultrasonography with spectral analysis following intracorporeal injection of a vasoactive drug. This method allows detection of venous leaks with a sensitivity of 55% to 100% and specificity of 69% to 88%. Arterial problems are also detected with a sensitivity of 82% to 100% and specificity of 64% to 96%. Other special tests, such cavernosometry, cavernosography, or penile arteriography, are rarely done in the evaluation of erectile dysfunction.

Treatment

If initiation of a medication is temporally related to the onset of erectile dysfunction, then the offending medication should be discontinued. Erectile dysfunction will typically return to baseline. However, if the onset of erectile dysfunction is long after the initiation of medication known to be associated with erectile dysfunction, then discontinuation of that medication will not usually improve erectile function. For psychogenic erectile dysfunction, simple reassurance and explanation, formal psychotherapy, and various forms of behavioral therapy have a reported 40% to 70% success rate.

For psychogenic or organic erectile dysfunction in a patient with a normal physical examination, most clinicians elect to begin with a therapeutic trial of an oral type 5 phosphodiesterase (PDE5) inhibitor that potentiates the effects of nitric oxide by inhibiting the breakdown of cyclic guanosine monophosphate. The initial oral dose is usually 50 mg of sildenafil (Viagra), 10 mg of tadalafil (Cialis), or 10 mg of vardenafil (Levitra). These medications should be tried only if the patient is not taking nitrates, has not had a myocardial infarction in the last 6 months, and does not have unstable angina, hypotension, severe congestive heart failure, or retinitis pigmentosa. These drugs are effective in 70% to 80% of patients with a wide variety of causes of erectile dysfunction, including psychogenic ones. Side effects include headache (16%) and visual disturbances (3%). They are absolutely contraindicated in patients taking nitrates for vascular disease or amyl nitrate that is sometimes taken to enhance sexual pleasure.

Vasoactive drugs including prostaglandin E_1, papaverine hydrochloride, and phentolamine mesylate (either alone or in combination), may induce an erection following intracavernous injection. Of these, the only Food and Drug Administration-approved agent is prostaglandin E_1 (alprostadil). This drug is titrated for each individual to the minimal effective dosage (typically 2.5 to 60 μg per injection). In clinical studies, up to 90% of men with erectile dysfunction developed erections with intracavernosal injections. Side effects include penile pain (up to 50% of injections), hematoma (1%-3%), and, less commonly, penile fibrosis (2%) and priapism (0.5%-1.0%). Intraurethral insertion of a 1.4-mm pellet containing alprostadil leads to satisfactory erections in 60% to 70% of patients, with tumescence beginning within 10 minutes and lasting 30 to 60 minutes. The major side effects are penile or urethral pain (10%), and dizziness (1%-2%). Intracavernosal and intraurethral alprostadil may be effective in patients with erectile dysfunction who have had no or an incomplete response to an oral PDE5 inhibitor. Alprostadil and oral PDE5 inhibitors also act synergistically when used in combination.

Devices have been developed that use suction to induce penile engorgement and constrictive bands to maintain the ensuing erection. Erections are achieved in 90% of patients with an approximately 70% couple satisfaction rate. Alternatively, a surgically implanted semi-rigid or inflatable penile prosthesis provides satisfactory results in 85% to 90% of cases, but these devices are used infrequently. Repair of venous leaks and microsurgical revascularization of arterial lesions have had variable success rates; vascular surgery is seldom used as a treatment for erectile dysfunction.

GYNECOMASTIA

Gynecomastia is common during the neonatal period and is present in about 70% of pubertal males (Chapter 15). Clinically apparent gynecomastia has been noted at autopsy in almost 1% of adult males, and 40% of autopsied males have histologic evidence of gynecomastia.

Etiology and Pathophysiology

The causes of gynecomastia are listed in Table 12–7. Several mechanisms have been proposed to account for this disorder. All involve a relative imbalance between estrogen and androgen concentrations or action at the mammary gland level. Peripubertal gynecomastia is considered to be a normal variant and is due to a relative imbalance in the production of estrogen relative to testosterone early in puberty. Decreased serum free testosterone concentrations due to primary gonadal disease or decreased androgen action in patients with the androgen resistance syndromes results in variable degrees of gynecomastia. Acute or chronic excessive stimulation of the Leydig cells by LH increases aromatization of testosterone to estradiol. This mechanism may be responsible for the more significant gynecomastia seen in hypergonadotropic states such as Klinefelter syndrome and adult Leydig cell failure. The rise of serum gonadotropins during puberty may lead to an estrogen-androgen imbalance by similar mechanisms. Patients who are malnourished or have systemic illness may develop gynecomastia during refeeding or treatment of the underlying disorder. Malnourishment and chronic illness are accompanied by a reduction in gonadotropin secretion, and during recovery, serum LH concentration rises and stimulates Leydig cell production of estrogens relative to testosterone.

Because hCG has LH-like activity, hCG-producing trophoblastic or nontrophoblastic tumors often result in relatively high serum estrogen to testosterone concentrations. In addition, some of these tumors are able to convert estrogen precursors into estradiol. Feminizing adrenocortical and Leydig cell neoplasms may directly secrete excessive quantities of estrogens. The mechanisms by which hyperprolactinemia produce gynecomastia are unclear, but the primary explanation appears be the known suppressive effect of hyperprolactinemia on gonadotropin secretion resulting in lower testosterone production and a lower testosterone:estrogen ratio.

Drugs that cause gynecomastia may reduce androgen synthesis (eg, spironolactone, ketoconazole), peripherally antagonize

TABLE 12–7   Causes of gynecomastia.

Physiologic
Neonatal
Pubertal
Aging

Drug Induced
Hormones
   Growth hormone (G)
   Human chorionic gonadotropin (G)
   Estrogens and estrogen agonists (G)
   Environmental estrogens (F)
Androgen synthesis inhibitors
   5-reductase inhibitors
      Dutasteride (G)
      Finasteride (G)
   Gonadotropin-releasing hormone analogs
      Goserelin (G)
      Leuprorelide (G)
Anti-androgens
   Spironolactone (G)
   Cyproterone (G)
   Biclutamide (G)
   Flutamide (G)
   Nilutamide (G)
Antibiotics and anti-viral drugs
   Anti-retroviral therapy (F)
   Ketoconazole (G)
Anti-ulcer medications
   Cimetidine (G)
   Omeprazole (F)
Cancer chemotherapeutic agents
   Alkylating agents (F)
Cardiovascular drugs
   Calcium channel blockers
      Nifedipine (F)
      Verapamil (F)
Drugs of abuse
   Alcohol (F)
   Opioids (F)
   Anabolic androgenic steroids (F)
Psychoactive agents
   Haloperidol (F)
   Phenothiazines (F)
   Second generation anti-psychotics (F)

Endocrine
Hypogonadism
Hyperprolactinemia
Hyperthyroidism
Androgen receptor disorders
Excessive aromatase activity

Systemic Diseases
Hepatic cirrhosis
Uremia
Recovery from malnourishment

Neoplasms
Testicular germ cell or Leydig cell tumors
Feminizing adrenocortical adenoma or carcinoma
hCG-secreting nontrophoblastic neoplasms

Idiopathic

The level of evidence supporting the drug cited as causing gynecomastia varies and is graded as follows.

G = good evidence of causal relationship based on a systematic review of randomized controlled trials, randomized placebo-controlled trial, or prospective cohort studies and a pathophysiological mechanism for gynecomastia.

F = fair evidence of causal relationship based on retrospective studies or case-control studies or case series and a pathophysiological mechanism for gynecomastia.

androgen action (spironolactone, cimetidine), or interact with breast estrogen receptors (spironolactone and phytoestrogens in marijuana). As part of the evaluation of recent onset of gynecomastia, the clinician should review all of the patient's prescription and non-prescription medications, creams, and lotions for potential causes.

Finally, it has been proposed that patients with idiopathic and familial gynecomastia have breast glandular tissue that is inordinately sensitive to normal circulating levels of estrogen or excessively converts estrogen precursors to estrogens.

Pathology

Three histologic patterns of gynecomastia have been recognized. The florid pattern consists of an increase in the number of budding ducts, proliferation of the ductal epithelium, periductal edema, and a cellular fibroblastic stroma. The fibrous type has dilated ducts, minimal duct epithelial proliferation, no periductal edema, and a virtually acellular fibrous stroma. An intermediate pattern contains features of both types.

The duration of gynecomastia is the most important factor in determining the pathologic picture. Approximately 75% of patients with gynecomastia of 4 months' duration or less exhibit the florid pattern, whereas 90% of patients with gynecomastia lasting a year or more have the fibrous type. Between 4 months and 1 year, 60% of patients have the intermediate pattern. The pathologic picture also correlates with the likelihood of resolution of gynecomastia; patients with more florid pathology are more likely to have reduction in breast tissue with treatment of the cause of the gynecomastia.

Clinical Features

A. Symptoms and signs   The principal complaint is unilateral or bilateral concentric enlargement of breast glandular tissue. Nipple or breast pain is present in one-fourth of patients and objective tenderness in about 40%. Breast tenderness and pain may be unilateral. Galactorrhea is very rare in men with gynecomastia. Histologic examination has demonstrated that gynecomastia is almost always bilateral, although it is often asymmetric and is sometimes nonpalpable on one side. Breast or nipple discomfort generally lasts less than 1 year. Other than cosmetic concerns, chronic gynecomastia is virtually always asymptomatic. Symptoms and signs of underlying disorders may be present.

The most important elements of the physical examination are breast and testicular examinations. The clinician must distinguish between pseudogynecomastia (fat tissue) and true gynecomastia. Pinching the tissue between thumb and index finger or between two hands allows the examiner to feel for an edge that represents the interface between normal tissue and rubbery glandular breast tissue. The examiner can often "flip the edge" of this interface so that the discoid breast tissue can be moved up and down or in and out of the plane of the surrounding tissue. Comparision of the consistency with fat tissue in the abdomen is useful. Gynecomastia is considered significant in adult men by some experts when there is more than or equal to 2 cm of palpable breast tissue.

The testicles should be examined for size, consistency, and the presence of a mass. A mass might represent a hormone-producing tumor. In addition to a careful examination of the breasts and testes, the clinician should examine for signs of hyperthyroidism, Cushing syndrome, and acromegaly.

B. Laboratory findings   The vast majority of adult men with nontender gynecomastia will not have an identifiable cause, and extensive laboratory testing is not useful. Breast tenderness suggests growth of breast tissue and recent onset of gynecomastia, and men with tender gynecomastia are more likely to have an endocrinopathy as the cause. Figure 12–8 depicts a rational diagnostic algorithm for the evaluation of nontender and tender gynecomastia.

Specific laboratory testing for liver, kidney or endocrine diseases associated with gynecomastia should be considered based on the history and examination. Measurement of testosterone, LH, and FSH on a serum sample obtained in the early morning is reasonable. Table 12–8 lists the pattern of testosterone, LH, and FSH concentrations seen in various causes of gynecomastia.

Differential Diagnosis

Gynecomastia should be differentiated from pseudogynecomastia due to obesity and from much rarer causes including lipomas, neurofibromas, or carcinoma of the breast. Breast lipomas, neurofibromas, and carcinoma are usually unilateral, painless, and
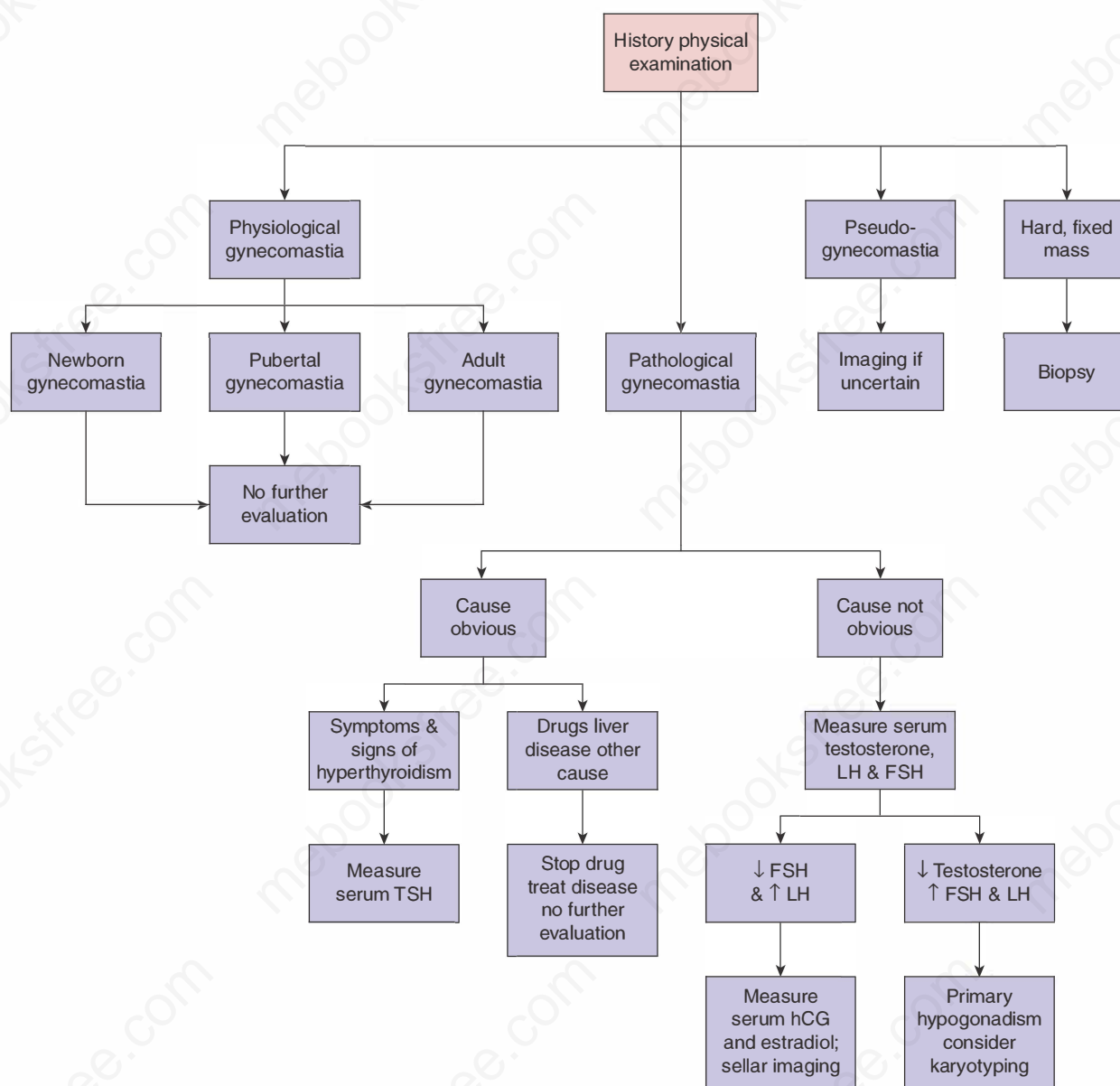
FIGURE 12–8   Diagnostic evaluation for endocrine causes of gynecomastia (hCG, human chorionic gonadotropin; LH, luteinizing hormone; FSH, follicle stimulating hormone; TSH, thyrotropic hormone).

TABLE 12–8 Typical hormone concentrations in the evaluation of gynecomastia.

Cause of Gynecomastia	Total Testosterone	LH	FSH	Other
Primary hypogonadism	↓	↑	↑	
Secondary hypogonadism	↓	Normal/↓	Normal/↓	
Hyperthyroidism	Normal/↑	↑	↑	↓/normal free testosterone
Hepatopathy	Normal/↑	Normal	Normal	Normal free testosterone
Exogenous hCG or hCG-producing tumor	↑	↓	↓	
Estradiol-producting testicular tumor	↓	↓	↓	Mass on scrotal ultrasound

eccentric, but gynecomastia characteristically begins in the sub-areolar areas and enlarges concentrically. Lipomas are usually easily recognized as mobile smooth nodules whereas male breast cancer often presents as a hard, fixed mass.

Complications and Sequelae

There are no complications other than emotional distress due to the cosmetic appearance. With the exception of patients with Klinefelter syndrome, individuals with gynecomastia do not have an increased risk of development of breast carcinoma.

Treatment

A. Medical treatment   The underlying disease should be corrected if possible, and offending drugs should be discontinued. Selective estrogen receptor modulators, such as tamoxifen or raloxifene, have been found useful in relieving pain and reversing gynecomastia in some patients. Although it would be expected that reduction of estradiol production by inhibition of aromatase would be effective for gynecomastia, the largest aromatase study to date demonstrated no benefit for anastrazole compared to placebo. Short-term administration of tamoxifen (10-20 mg daily for 3-6 months) may be useful in boys with significantly symptomatic pubertal gynecomastia or men with recent onset of very symptomatic idiopathic gynecomastia. Tamoxifen also effectively prevents the development of gynecomastia in many men who are starting therapy for prostate cancer with antiandrogens. Medical therapy is ineffective for chronic, fibrous gynecomastia.

B. Surgical treatment   Reduction mammoplasty should be considered for cosmetic reasons in any patient with longstanding gynecomastia that is in the fibrotic stage. Causes of gynecomastia should be addressed prior to reduction mammoplasty.

C. Radiation therapy   Low-dosage external beam radiation therapy (900 cGy or less) is effective, but less effective than tamoxifen, for prevention of gynecomastia due to anti-androgen monotherapy for prostate cancer.

Course and Prognosis

Pubertal gynecomastia usually regresses spontaneously within 1 or 2 years. Patients who develop drug-induced gynecomastia generally have complete or near-complete regression of the breast changes if the offending drug is discontinued during the early, florid stage. Once gynecomastia from any cause has reached the fibrotic stage, little or no spontaneous regression occurs, and medical therapy is ineffective.

TESTICULAR TUMORS

Testicular neoplasms account for 1% to 2% of all male-specific malignant neoplasms and 4% to 10% of all genitourinary neoplasms. They are the most frequent type of cancer in men between 20 and 34 years of age. The incidence is 5.4 per 100,000 men in the United States and 4.6 to 7.3 per 100,000 men in Europe. The incidence is lower in nonwhite than in white populations. Ninety-five percent of testicular tumors are of germ cell origin; 5% are composed of stromal or Leydig cell neoplasms.

Etiology and Pathophysiology

The cause of germ cell testicular tumors is not known. Predisposing factors include testicular maldescent, Klinefelter syndrome, Down syndrome, and tobacco use. Exposure to environmental compounds with estrogenic activity has been associated with increased risk of testicular cancer. Although trauma is frequently cited as an etiologic factor in testicular tumors, no causal relationship has been established.

Pathology

A. Germ cell tumors   Seminomas account for 33% to 50% of all germ cell tumors. They are composed of round cells with abundant cytoplasm, prominent nuclei, and large nucleoli. The cells are arranged in cords and nests and have a thin delicate network of stromal connective tissue. Embryonal cell neoplasms comprise 20% to 33% of germ cell tumors. These tumors have multiple histologic patterns composed of cuboidal pleomorphic cells. One distinct pattern of cellular arrangement is the endodermal sinus tumor (yolk sac tumor), the most frequent germ cell neoplasm found in infants. Immunohistochemical techniques have localized alpha-fetoprotein to the embryonal cells. About 10% of germ cell tumors are teratomas, which are composed of well-differentiated cells derived from all three germ layers.

When one or more of the teratoid elements are malignant or are mixed with embryonal carcinoma cells, the term teratocarcinoma is applied. These tumors account for one-tenth to one-third of germ cell neoplasms. Choriocarcinoma is the rarest form of germ cell tumor (2%) and is composed of masses of large, polymorphic, multinucleated syncytiotrophoblastic cells. Although pure choriocarcinoma is rare, many testicular tumors contain an occasional trophoblastic giant cell. Immunohistochemical techniques have shown that these cells are the source of hCG in such tumors. Germ cell tumor sometimes produce enough hCG to cause bilateral gynecomastia.

B. Leydig and Sertoli cell tumors   Leydig cell (interstitial cell) tumors are rare. Most are benign and are composed of sheets of oval to polygonal cells arranged in lobules separated from one another by thin strands of connective tissue. Malignant Leydig cell tumor disseminates by both lymphatic and venous channels, with initial metastatic deposits being found in the regional lymph nodes, followed by metastases to liver, lung, and bone. Leydig cell tumors may produce testosterone or estrogen. Sertoli cell tumors also are rare, are generally benign, and are composed of large tubules, a thick basement membrane with enlarged cuboidal Sertoli cells. There may be extensive calcification within the tumors and the tumors may be multifocal and bilateral. Sertoli cell tumors may produce estrogen.

Clinical Features

A. Symptoms and signs

1. **Germ cell tumors**—Testicular tumors usually present as painless enlargement of a testicle. Thus, about 80% of patients note a testicular swelling or mass, whereas only 25% complain of testicular pain or tenderness. About 6% to 25% of patients give a history of testicular trauma that led to the discovery of the testicular mass. About 5% to 10% of patients present with symptoms of distant metastatic disease, including backache, skeletal pains, gastrointestinal symptoms and abdominal pains, inguinal adenopathy, cough, hemoptysis, and neurological dysfunction.

    A testicular mass or generalized enlargement of the testis is often present on examination. In 5% to 10% of patients, a coexisting hydrocele may be present. In the presence of metastatic disease, supraclavicular and retroperitoneal lymph node enlargement may be present.

2. **Leydig and Sertoli cell tumors**—In children, Leydig cell tumors of the testes may produce sexual precocity, with rapid skeletal growth and development of secondary sexual characteristics. Adults with Leydig cell tumors usually present with a testicular mass and occasionally with tender gynecomastia. Sertoli cell tumors may be associated with gynecomastia and feminization due to excessive aromatase activity. Sertoli tumors are associated with the Peutz-Jeghers syndrome (gastrointestinal polyposis and oval, irregularly pigmented lip macules) and the Carney complex (cardiac myxomas, spotty cutaneous pigmentation, primary pigmented nodular adrenocortical disease with hypercortisolism).

B. Laboratory findings

1. **Germ cell tumors**—The tumor markers hCG, alpha-fetoprotein, and lactate dehydrogenase (LDH) should be measured in every male presenting with a testicular mass. hCG is found in the sera of 5% to 10% of males with seminoma, over half of patients with teratocarcinoma or embryonal cell carcinoma, and all patients with choriocarcinoma. hCG should be measured as the beta subunit or other hCG-specific immunoassay method. Elevated serum immunoreactive alpha-fetoprotein concentrations are found in almost 70% of patients with nonseminomatous forms of germ cell neoplasms. Both markers are elevated in over 50% of patients with nonseminomatous germ cell tumors, and at least one of the markers is elevated in 85% of such patients. These markers can also be used to monitor the results of therapy. Serum LDH concentrations reflect tumor growth rate and cellular proliferation, and high serum concentrations are associated with a poor prognosis. Serum LDH is increased in about 80% of patients with advanced seminomas and 60% of advanced nonseminomatous tumors.

2. **Leydig and Sertoli cell tumors**—Serum DHEA sulfate and estrogen concentrations may be normal or increased, while serum testosterone concentrations tend to be low or within the normal adult range in patients with Leydig cell tumors. Individuals with Sertoli cell tumors may have elevated serum estradiol concentrations.

C. Imaging studies   Scrotal ultrasonography is useful for the evaluation of any testicular mass and may be required to visualize small testicular tumors. Staging of testicular tumors requires chest and abdominal CT scans and other radiologic procedures depending on the type of tumor and the symptoms.

Differential Diagnosis

Testicular tumors are sometimes misdiagnosed as epididymitis or epididymo-orchitis. An inflammatory reaction of the epididymis often involves the vas deferens. Therefore, both the vas and the epididymis are thickened and tender on examination during the acute disease. Because hydrocele may coexist with testicular tumor, all patients with large hydroceles (large enough to make the testicular examination difficult) should have a scrotal ultrasound.

Other conditions in the differential diagnosis include inguinal hernia, hematocele, hematoma, torsion, spermatocele, varicocele, and (rarely) sarcoidosis, tuberculosis, and syphilitic gumma. Ultrasonic examination of the scrotum may help distinguish between testicular tumors and extratesticular disease such as acute or chronic epididymitis, spermatocele, or hydrocele.

Benign Leydig cell tumors of the testes must be differentiated from adrenal rest tumors in patients with congenital adrenal hyperplasia. Because the testes and the adrenals are derived from the same embryologic source, ectopic adrenal tissue may be found to migrate with the testes. Testicular adrenal rests enlarge under the stimulus of higher serum ACTH concentrations in patients with congenital adrenal hyperplasia or Cushing disease. Adrenal

rests tend to be bilateral, whereas Leydig cell tumors are generally unilateral. Both may be associated with elevated serum DHEA sulfate and estradiol concentrations.

Treatment

A. Germ cell tumors   Seminomas are quite radiosensitive, and disease localized to the testes is usually treated with radical inguinal orchidectomy and 2000 to 4000 cGy of conventional radiotherapy, delivered to the ipsilateral inguinal-iliac and bilateral para-aortic lymph nodes to the level of the diaphragm. For disease that has spread to the lymph nodes below the diaphragm, additional whole abdominal radiotherapy and prophylactic mediastinal and supraclavicular lymph node irradiation are usually given. Widely disseminated disease is generally treated with a combination of radiotherapy and chemotherapy, especially with bleomycin, etoposide, and cisplatin.

Nonseminomatous tumors are treated with orchidectomy, retroperitoneal lymph node dissection, and, if necessary, radiotherapy or chemotherapy (or both). Although many chemotherapeutic agents have been used, combinations of etoposide, bleomycin, and cisplatin currently appear to produce the best overall results. Patients with nonseminomatous tumors treated by these means should be monitored with serial measurements of serum hCG and alpha-fetoprotein.

B. Leydig and Sertoli cell tumors   Leydig and Sertoli cell tumors of the testes are treated by unilateral radical inguinal orchidectomy. Objective remissions of malignant Leydig cell tumors have been noted following treatment with mitotane.

Course and Prognosis

A. Germ cell tumors   In patients with seminoma confined to the testicle, the 5-year survival rates after orchidectomy and radiotherapy are 98% to 100%. Disease in the lymph nodes below the diaphragm also has an excellent prognosis, with 5-year survival rates of 80% to 85%. Disease above the diaphragm and disseminated disease have overall 5-year survival rates of about 70%.

In patients with nonseminomatous germ cell tumors, aggressive surgery and combination chemotherapy have raised the 5-year survival rates from less than 20% to 60% to more than 95%.

B. Leydig and Sertoli cell tumor   Removal of a benign Leydig or Sertoli cell tumor is accompanied by regression of iso- or heterosexual precocious puberty in children or feminization in adults. The prognosis for malignant Leydig cell tumor is poor, with most patients surviving less than 2 years from the time of diagnosis.

REFERENCES

Physiology

Amory JK, Bremner WJ. Regulation of testicular function in men: implications for male hormonal contraceptive development. *J Steroid Biochem Mol Biol.* 2003;85:357. [PMID: 12943722]

Finkelstein JS, Lee H, Burnett-Bowie SM, et al. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med.* 2013;369:1011. [PMID: 24350954]

Josso N, Picard JY, Rey R, Clemente N. Testicular anti-Mullerian hormone: history, genetics, regulation and clinical application. *Pediat Endocrinol Rev.* 2006;3:347. [PMID: 16816803]

O'Hara L, Smith LB. Androgen receptor roles in spermatogenesis and infertility. *Best Pract Res Clin Endocrinol Metab.* 2015;29:595. [PMID: 26303086]

Rajender S, Singh L, Thangaraj K. Phenotypic heterogeneity of mutations in androgen receptor. *Asian J Androl.* 2007;9:147. [PMID: 17334586]

Sofikitis N, Giotitsas N, Tsounapi P, Baltogiannis D, Giannakis D, Pardalidis N. Hormonal regulation of spermatogenesis and spermiogenesis. *J Steroid Biochem Molecular Biol.* 2008;109:323. [PMID: 18400489]

Androgen Therapy

Berkseth KE, Thirumalai A, Amory JK. Pharmacologic therapy in men's health: hypogonadism, erectile dysfunction, and benign prostatic hyperplasia. *Med Clin North Am.* 2016;100:791. [PMID: 27235615]

Hypogonadism

Anawalt BD. The Model T. *J Clin Endocrinol Metab.* 2016;101:2640-2. [PMID: 27381959]

Bhasin S, Cunningham GR, Hayes FJ, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society practice guideline. *J Clin Endocrinol Metab.* 2010;95:2536. [PMID: 20525905]

Cunningham GR, Stephens-Shields AJ, Rosen RC, et al. Testosterone treatment and sexual function in older men with low testosterone levels. *J Clin Endocrinol Metab.* 2016;101:3096. [PMID: 27355400]

Kolon TF, Herndon CD, Baker LA, et al. Evaluation and treatment of cryptorchidism: AUA Guideline. *J Urol.* 2014;192:337. [PMID: 24857650]

Ranum LPW, Day JW. Myotonic dystrophy: RNA pathogenesis comes into focus. *Am J Hum Genet.* 2004;74:793. [PMID: 15065017]

Snyder PJ, Bhasin S, Cunningham GR, et al. Effects of testosterone treatment in older men. *N Engl J Med.* 2016;374:611. [PMID: 26886521]

Van der Burgt I. Noonan syndrome. *Orphanet J Rare Dis.* 2007;2:4. [PMID: 17222357]

Virtanen HE and Toppari J. Cryptorchidism and fertility. *Endocrinol Metab Clin N Am.* 2015;44:751. [PMID: 26568490]

Wang C, Nieschlag E, Swerdloff R, et al. ISA, ISSAM, EAU, EAA and ASA recommendations: investigation, treatment and monitoring of late-onset hypogonadism in males. *J Androl.* 2009;30:1. [PMID: 18772485]

Wu FCW, Tajar A, Beynon JM, et al. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med.* 2010;363:123. [PMID: 20554979]

Infertility

Anawalt BD. Approach to male infertility and induction of spermatogenesis. *J Clin Endocrinol Metab.* 2013;98:3532. [PMID: 24014811]

Baazem A, Belzile E, Ciampi A, et al. Varicocele and male factor infertility treatment: a new meta-analysis and review of the role of varicocele repair. *Eur Urol.* 2011;60:796. [PMID: 21733620]

Evers JH, Collins J, Clarke J. Surgery or embolization for varicoceles in subfertile men. *Cochrane Database Syst Rev.* 2009;1:CD000479. [PMID: 19160180]

Fullerton G, Hamilton M, Maheshwari A. Should non-mosaic Klinefelter syndrome men be labelled as infertile in 2009. *Hum Reprod.* 2010;25:588. [PMID: 20085911]

Jefferys A, Siassakos D, Wardle P. The management of retrograde ejaculation: a systematic review and update. *Fertil Steril.* 2012;97:306. [PMID: 22177462]

McLachlan RI, O'Bryan MK. State of the art for genetic testing of infertile men. *J Clin Endocrinol Metab.* 2010;95:1013. [PMID: 22998070]

Practice Committee of the American Society for Reproductive Medicine in collaboration with the Society for Male Reproduction and Urology. The management of infertility due to obstructive azoospermia. *Fertil Steril.* 2008;90(Suppl 3):S121. [PMID: 19007607]

Sunderam S, Chang J, Flowers L, et al. Centers for Disease Control and Prevention (CDC). Assisted reproductive technology surveillance—United States, 2006. *MMWR Surveill Summ.* 2009;58:1.

Erectile Dysfunction

Bella AJ, Lee JC, Carrier S, Benard F, Brock GB. 2015 CUA Practice guidelines for erectile dysfunction. *Can Urol Assoc J.* 2015;9:23. [PMID: 25737750]

Muneer A, Kalsi J, Nazareth I, Arya M. Erectile Dysfunction. *BMJ.* 2014;348:129. [PMID: 24468580]

Miner M, Seftel AD, Nehra A, et al. Prognostic utility of erectile dysfunction for cardiovascular disease in younger men and those with diabetes. *Am Heart J.* 2012;164:21. [PMID: 22795278]

Nehra A, Jackson G, Miner M, et al. The Princeton III Consensus recommendations for the management of erectile dysfunction and cardiovascular disease. *Mayo Clin Proc.* 2012;87:766. [PMID: 22862865]

Spitzer M, Basaria S, Travison TG, et al. Effect of testosterone replacement on response to sildenafil citrate in men with erectile dysfunction: a parallel, randomized trial. *Ann Intern Med.* 2012;157:681. [PMID: 23494931]

Gynecomastia

Braunstein GD. Clinical practice. Gynecomastia. *N Engl J Med.* 2007;357:1229. [PMID: 17881754]

Brinton LA. Breast cancer risk among patients with Klinefelter syndrome. *Acta Paediatr.* 2011;100:814. [PMID: 21241366]

Deepinder F, Braunstein GD. Drug-induced gynecomastia: an evidence-based review. *Expert Opin Drug Saf.* 2012;11:779. [PMID: 22862307]

Lapid O, van Wingerden J, Perlemuter L. Tamoxifen therapy for the management of pubertal gynaecomastia: a systematic review. *J Pediatr Endocrinol Metab.* 2013;26:803. [PMID: 23729603]

Niewoehner CB, Schorer AE. Gynaecomastia and breast cancer in men. *BMJ.* 2008;336:709. [PMID: 18369226]

Plourde PV, Reiter EO, Jou HC, et al. Safety and efficacy of anastrozole for the treatment of pubertal gynecomastia: a randomized, double-blind, placebo-controlled trial. *J Clin Endocrinol Metab.* 2004;89:4428. [PMID: 15356042]

Testicular Cancer

Cole LA, Kohorn EI. The need for an hCG assay that appropriately detects trophoblastic disease and other hCG-producing cancers. *J Reprod Med.* 2006;51:793.

Carmignani L, Salvioni R, Gadda F, et al. Long-term followup and clinical characteristics of testicular Leydig cell tumor: experience with 24 cases. *J Urol.* 2006;176:2040. [PMID: 17070249]

Feldman DR. Update in germ cell tumours. *Curr Opin Oncol.* 2015;27:177. [PMID: 25850580]

Female Reproductive Endocrinology and Infertility

Mitchell P. Rosen, MD and Marcelle I. Cedars, MD

ACTH	Adrenocorticotropic hormone	**IUI**	Intrauterine insemination
AFC	Antral follicle count	**LDL**	Low-density lipoprotein
AIRE	Autoimmune regulator gene	**LH**	Luteinizing hormone
AIS	Androgen insensitivity syndrome	**MBH**	Medial basal hypothalamus
AMH	Anti-Müllerian hormone	**MCR**	Metabolic clearance rate
APS	Autoimmune polyglandular syndrome	**MMP**	Matrix metalloproteinase
ArKO	Aromatase knockout	**MPA**	Medroxyprogesterone acetate
ART	Assisted reproductive therapy	**MT**	Menopausal transition
BBT	Basal body temperature	**OMI**	Oocyte maturation inhibitor
BMD	Bone mineral density	**PCOS**	Polycystic ovarian syndrome
BMP	Bone morphogenic protein	**POF**	Premature ovarian failure
cAMP	Cyclic adenosine monophosphate	**PR**	Production rate
CEE	Conjugated equine estrogen	**PRL**	Prolactin
CRH	Corticotropin-releasing hormone	**SERM**	Selective estrogen receptor modulator
DHEAS	Dehydroepiandrosterone sulfate	**SF-1**	Steroidogenic factor-1
DHT	Dihydrotestosterone	**SHBG**	Sex hormone–binding globulin
DXA	Dual-energy x-ray absorptiometry	**SNRI**	Serotonin-norepinephrine reuptake inhibitor
ERA	Estrogen Replacement and Atherosclerosis Trial	**SR**	Secretion rate
FGFR1	Fibroblast growth factor receptor 1	**SRY**	Sex-determining region of the Y gene
FMR1	Fragile X gene	**SSRI**	Selective serotonin reuptake inhibitor
FSH	Follicle-stimulating hormone	**StAR**	Steroidogenic acute regulatory protein
FXTAS	Fragile-X tremor ataxia syndrome	**STRAW**	Stages of Reproductive Aging Workshop
GH	Growth hormone	**SWAN**	Study of Women's Health Across the Nation
GnRH	Gonadotropin-releasing hormone	**Tg**	Thyroglobulin
hCG	Human chorionic gonadotropin	**TGF**	Transforming growth factor
HDL	High-density lipoprotein	**TPO**	Thyroid peroxidase
HERS	Heart and Estrogen-Progestin Replacement Study	**TRH**	Thyrotropin-releasing hormone
HPO	Hypothalamic pituitary ovarian	**TSH**	Thyroid-stimulating hormone
HSD	Hydroxysteroid dehydrogenase	**VEGF**	Vascular endothelial growth factor
HRT	Hormone replacement therapy	**WHI**	Women's Health Initiative
HT	Hormone therapy		
IGF-BP	Insulin-like growth factor binding protein		

EMBRYOLOGY AND ANATOMY

No gene has yet been identified that generates an ovary from an undifferentiated gonad. It is only in the absence of the sex-determining region of the Y gene (*SRY*) that the gonad develops into an ovary. (For more details, see the discussion of sexual differentiation in Chapter 14.) Primordial germ cells, which give rise to oocytes or spermatogonia, are first identifiable in the yolk sac endoderm (hindgut) at 3 to 4 weeks of gestation. Once specified, they migrate and proliferate en route through the dorsal mesentery into the gonadal ridge, which is located lateral to the dorsal mesentery of the gut and medial to the mesonephros (Figure 13–1). Studies in mice have suggested that the process of proliferation and navigation to the gonad depends on several genes, including *Steel* (kit ligand and receptor), β_1 *integrin, pog* (proliferation of germ cells), and many cytokines. Failure of primordial germ cells to develop or migrate into the gonadal ridge results in failure of ovarian development. In contrast, it is suggested that male gonadal development may continue to develop into functional testis despite the absence of germ cells.

The germ cells that reach the gonadal ridge (6 weeks of gestation) continue to proliferate and are referred to as oogonia (premeiotic germ cells). At 10 to 12 weeks of gestation, some oogonia leave the mitotic pool and begin meiosis, where they arrest in prophase I (dictyotene stage). These arrested germ cells are now called primary oocytes. By 16 weeks, primordial follicles are first identified, making a clear distinction for gonadal differentiation into an ovary. At approximately 20 weeks of gestation, a peak of 6 to 7 million germ cells (two-thirds of them primary oocytes and one-third oogonia) are present in the ovaries. During the second half of gestation, the rate of mitosis rapidly decreases, and the rate of oogonial and follicular atresia increases. Those oogonia that are not transformed into primary oocytes will undergo atresia before birth. This leads to a reduction in the number of germ cells, resulting in a total of 1 to 2 million germ cells at birth. No germ cell mitosis occurs after birth, whereas follicular atresia continues, with the result that the average girl entering puberty has only 300,000 to 400,000 germ cells.

The ovary is organized into an outer cortex and an inner medulla. The germ cells are located within the cortex. Along the outer surface of the cortex is the germinal epithelium. This cell layer is composed of cuboidal cells resting on a basement membrane and forms a continuous layer with the peritoneum. Even though it is called the germinal epithelium, there are no germ cells within this layer. During embryonic development, the epithelial cells proliferate and enter the underlying tissue of the ovary to form cortical cords. When the primordial germ cells arrive at the genital ridge, they are incorporated into these cortical cords. At the same time the germ cells migrate from the yolk sac, the stromal cells of the ovary (granulosa and interstitial cells) migrate from the mesonephric tubules into the gonad. **Primordial follicles** form within the cortical cords. They are composed of a primary oocyte and one layer of granulosa cells with its basement membrane. Oocytes not surrounded by granulosa cells are lost, probably by apoptosis. This finite follicle population represents the pool of germ cells that will ultimately be available to enter the follicular cycle.

During fetal development, the gonad is held in place by the suspensory ligament at the upper pole and the gubernaculum at the lower pole. The final location of the gonad is dependent on hormone production. In the presence of testosterone, the gubernaculum grows while the suspensory ligament regresses. As the gubernaculum continues to grow, the gonad (testis) descends into the scrotum. In contrast, when testosterone is absent, the suspensory ligament remains, and the gubernaculum regresses. This process maintains the gonad (ovary) in the pelvis.

The remainders of the female internal reproductive organs are formed from the paramesonephric (Müllerian) ducts. In the absence of anti-Müllerian hormone (AMH), the paramesonephric system develops into the uterine (fallopian) tubes, uterus, cervix, and upper third of the vagina. Unlike Wolffian duct differentiation, the development of the female reproductive tract is not
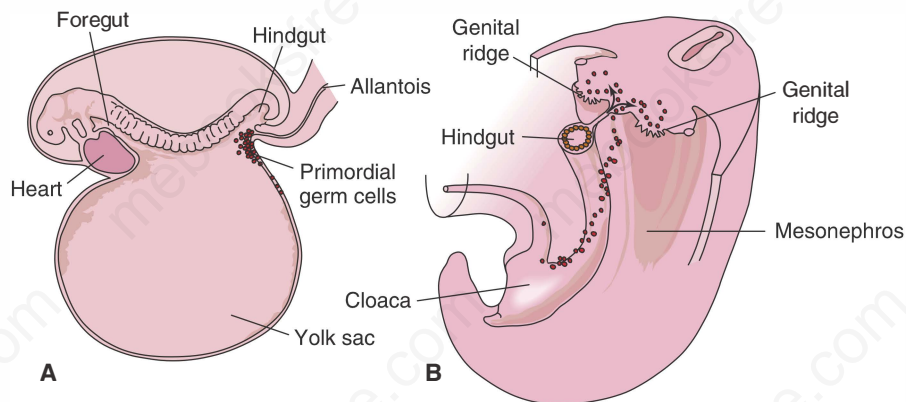
FIGURE 13–1   A. Schematic drawing of a 3-week-old embryo showing the primordial germ cells in the wall of the yolk sac, close to the attachment of the allantois. **B.** Drawing to show the migrational path of the primordial germ cells along the wall of the hindgut and the dorsal mesentery into the genital ridge.
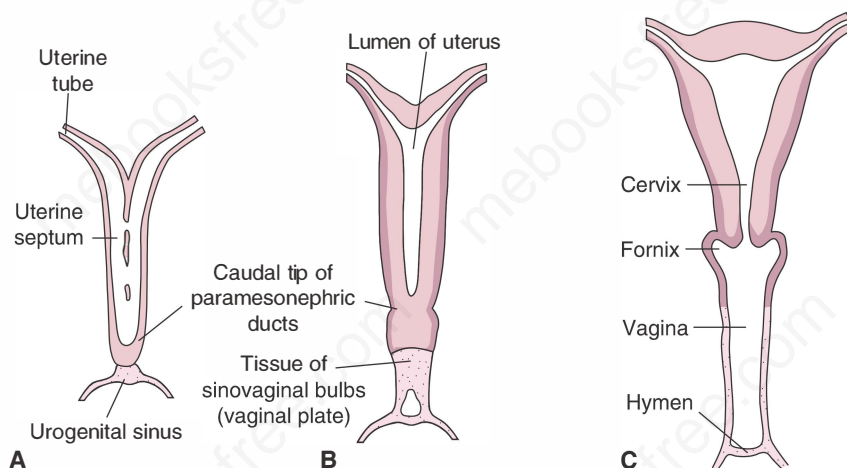
FIGURE 13–2 Schematic drawing showing the formation of the uterus and vagina. **A.** At 9 weeks. Note the disappearance of the uterine septum. **B.** At the end of the third month. Note the tissue of the sinovaginal bulbs. **C.** Newborn. The upper portion of the vagina and the fornices are formed by vacuolization of the paramesonephric tissue and the lower portion by vacuolization of the sinovaginal bulbs. (Reproduced with permission from Langman J, Sadler TW. *Langman's Medical Embryology.* 8th ed. Philadelphia: Lippincott Williams & Wilkins; 2000.)

dependent on hormone production. (For more details, see "Human Sex Differentiation" in Chapter 14.) Briefly, the Müllerian buds are formed lateral to the Wolffian ducts and the gonadal ridge after 37 days of gestation (Figure 13–2). These buds elongate, extend caudally, and cross medially to the Wolffian ducts by 8 weeks of gestation. By 10 weeks, the adjacent paired Müllerian ducts meet in the midline as they join the Müllerian tubercle (urogenital sinus). Over the next couple of weeks, the Wolffian ducts degenerate, and the paired Müllerian ducts fuse and begin to canalize. The intervening septum resorbs between the 12th and 16th weeks of gestation, resulting in a single uterine cavity. The most cranial parts of the Müllerian ducts remain unfused and form the uterine tubes, which remain patent with the coelom (future peritoneal cavity). The caudal segments stimulate solid cords to extend from the Müllerian tubercle to the sinovaginal bulbs from the posterior aspect of the urogenital sinus. In turn, the sinovaginal bulbs extend cranially and fuse with the vaginal cords, forming the vaginal plate. The vagina is subsequently formed by canalization of the vaginal plate. It is ultimately the cervix and the upper one-third of the vagina that are derived from the Müllerian structures, and the remaining lower two-thirds are formed from the urogenital sinus.

The uterus is composed of endometrium (innermost lining), myometrium, and serosa. The adult uterus is a pear-shaped hollow organ. The cervical portion extends approximately 2 cm into the vagina, and the remaining corpus extends approximately 6 cm into the abdomen. The normal adult uterus weighs 40 to 80 g.

The uterus is located in the pelvis and rests on the pelvic floor. From 70% to 80% of the time, the uterine position is anteflexed (cervical-uterine corpus angle) and anteverted (cervical-vaginal angle). Therefore, when a woman is standing, the corpus of the uterus is horizontal and resting on top of the bladder. The uterus has several paired ligaments that develop from thickenings of the peritoneum and serve to maintain this anatomic position. The cardinal ligament (Mackenrodt) is the main supporting ligament. It attaches to the lateral margins of the cervix at the upper vagina and extends to the lateral pelvic wall. The remaining ligaments—uterosacral, round, and broad—have a lesser role in supporting the uterus.

The uterine artery originates from the anterior division of the internal iliac artery (hypogastric), enters the cardinal ligament, and supplies the uterus. The uterine artery divides into a descending branch and an ascending branch known as the vaginal and arcuate arteries, respectively. The arcuate arteries anastomose with each other and form a vascular network around the uterus. The radial arteries branch off from the arcuate network and penetrate the uterus to supply the myometrium. Smaller basal branches and spiral arteries supply the endometrium.

The ovary is suspended in the pelvis and has three associated ligaments. The average adult ovary is 2.5 to 5 cm × 2.5 cm × 1 cm in size and weighs 3 to 8 g. The position of the ovary is variable, but in a nulliparous woman it is often located in a peritoneal depression on the pelvic sidewalls between the ureter and external iliac vein. The suspensory (infundibulopelvic) ligament attaches to the cranial pole of the ovary and extends to the pelvic brim. This ligament suspends the ovary in the pelvis and contains the ovarian vessels, lymphatics, and nerves. The utero-ovarian ligament attaches to the inferior pole of the ovary and extends to the uterus. The mesovarium connects the anterior portion of the ovary to the posterior leaf of the broad ligament. The blood supply of the ovary originates from the abdominal aorta, passes through the suspensory ligament, and enters the mesovarium to form an anastomotic network with branches from the uterine artery. The ovarian artery enters the ovarian hilum and branches into spiral arteries that enter the medulla and extend to the ovarian cortex. Other branches from the anastomotic network, located in the mesovarium, supply the uterine tubes.

OVARIAN STEROIDOGENESIS

The ovaries are not only the store for germ cells—they also produce and secrete hormones that are vital for reproduction and the development of secondary sexual characteristics. The next section will briefly discuss the biosynthesis of ovarian hormones.

In the ovary, the major source of hormone production is the maturing follicle. The components of the follicle are the theca cells, the granulosa cells, and the primary oocyte. The theca cells produce androgens, and the granulosa cells produce estrogens. The other stromal cells that contribute to androgen production can be divided into two populations of cells: the secondary interstitial cells (derived from theca) and the hilum cells. These cells are the major ones involved in ovarian hormone production during menopause (see later).

The ovarian hormones are derived from cholesterol. Steroidogenic cells acquire the cholesterol substrate from one of three sources. The most common source is plasma lipoprotein-carrying cholesterol, primarily in the form of low-density lipoprotein (LDL). Other minor sources include *de novo* synthesis from acetate and liberation from stored lipid droplets (cholesterol esters). Stimulation of ovarian cells by trophic hormones such as follicle-stimulating hormone (FSH) and luteinizing hormone (LH) facilitate uptake of cholesterol by increasing the number of LDL receptors on the cell surface. The LDL particle is subsequently internalized and degraded in the lysosome. The free cholesterol that is liberated from the lysosome is delivered to the mitochondria by an unknown mechanism, possibly via microfilaments and microtubules. The cholesterol is then translocated into the mitochondria by the steroidogenic acute regulatory protein (StAR). Thus, StAR is the rate-limiting step regulating substrate availability for steroidogenesis.

Although acute alterations in steroid production result from changes in delivery of cholesterol to the mitochondria, the long-term control of steroid synthesis results from regulation of gene expression. Most of the genes involved in steroidogenesis contain at least one steroidogenic factor-1 (SF-1) response element in the promoter region. These elements are critical for the regulation of steroidogenic genes as well as the development of adrenal gland, ovary, and testis. The importance of SF-1 for steroidogenesis is highlighted in knockout mice deficient in this transcription factor which lack adrenal glands and gonads. Although SF-1 is essential, the specific expression of each gene involves many other transcription factors that work independently or in concert with SF-1.

The rate-determining step that commits cholesterol to steroid synthesis is the cholesterol side chain cleavage enzyme reaction (P450scc) (Figure 13–3). This reaction converts cholesterol to pregnenolone, the precursor of steroid hormones, and takes place in the mitochondria. Pregnenolone is transported out of the
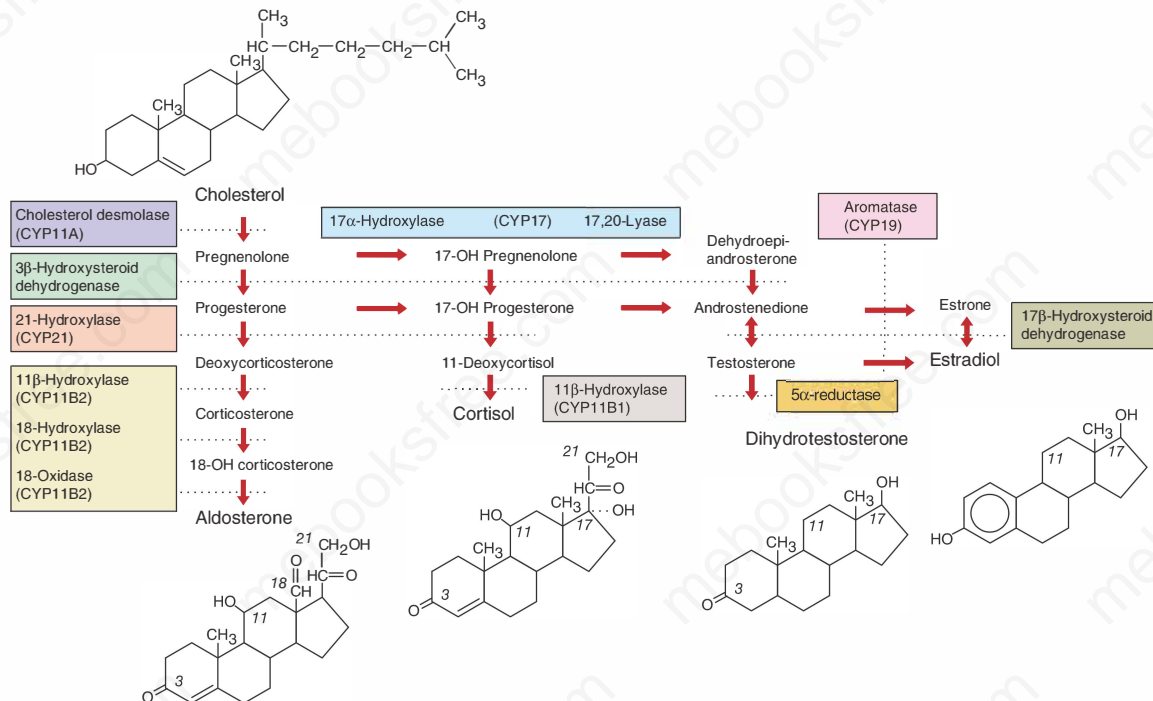
FIGURE 13–3 Pathways of steroid biosynthesis. The pathways for synthesis of progesterone and mineralocorticoids (aldosterone), glucocorticoids (cortisol), androgens (testosterone and dihydrotestosterone), and estrogens (estradiol) are arranged left to right. The enzymatic activities catalyzing each bioconversion are written in the boxes. For those activities mediated by specific cytochrome P450, the systematic name of the enzyme (*CYP* followed by a number) is listed in parentheses. CYPB2 and CYP17 have multiple activities. The planar structures of cholesterol, aldosterone, cortisol, dihydrotestosterone, and estradiol are placed near the corresponding labels. (Reproduced with permission from White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr Rev*. 2000 Jun;21(3):245-291.)

mitochondrion, and the remaining steps in sex steroid production take place primarily within the smooth endoplasmic reticulum (SER).

Once pregnenolone is formed, the particular hormones that are synthesized are dependent on the endocrine organ and cell type. For example, the main sources of sex steroids in the female come from the adrenal gland, ovary, and the periphery. The specific type of hormone synthesized is dependent on the specific gene expression within each cell type.

In the adrenal gland, there are three zones: zona glomerulosa, zona fasciculata, and zona reticularis. The cells in the different zones start with the same hormone precursor but differ in their secretory products. The glomerulosa produces mainly aldosterone, whereas cortisol and androgen are produced by the zona fasciculata and zona reticularis, respectively. The major androgen produced by the adrenal gland is dehydroepiandrosterone sulfate (DHEAS). Differences in enzymatic activity among cells in the various zones are what regulate hormone production. The zona reticularis and zona fasciculata lack 18 hydroxylase/oxidase (11β2-hydroxylase) activity, which is necessary for aldosterone synthesis (see Figure 13–3; see also Chapters 9 and 10). The zona glomerulosa lacks 17-hydroxylase and 17,20-lyase (CYP17), which are necessary for sex steroid synthesis.

Ovarian cells similarly secrete different hormones due to differential enzyme activity. The theca interstitial and secondary interstitial cells lack aromatase and hence are the androgen producers in the ovarian cortex. The granulosa cells, on the other hand, lack CYP17 and, therefore, secrete estrogens—mainly estradiol—from thecal cell-derived androgens in the proliferative phase and progesterone in the luteal phase (see later).

Adipose and skin make significant contributions to plasma concentrations of some sex steroids. Adipose tissue is able to sequester most steroids due to their lipid solubility. Fat also expresses genes capable of sex steroid metabolism (ie, aromatase). Skin significantly contributes to the plasma concentration of testosterone by utilizing circulating DHEAS and androstenedione as precursors.

The major circulating androgens include DHEAS, androstenedione, and testosterone. During the reproductive years, the ovaries are directly responsible for one-third of the testosterone production. The remaining two-thirds come from the periphery (17β-hydroxysteroid dehydrogenase [HSD] types 3 and 5) and is derived from ovarian and adrenal gland precursors—notably androstenedione, which is produced in equal proportions by the adrenal gland and the ovary. The adrenal gland may directly secrete testosterone, but its major contribution is derived from its production of precursors. Therefore, the ovaries are responsible for nearly two-thirds of the circulating testosterone. This differs from males, in whom only 5% of the circulating testosterone is derived from peripheral conversion of androstenedione. It is also estimated that more than 60% of the most potent androgen, dihydrotestosterone (DHT) in women, is produced in the skin (5α-reductase type 1 and 2) and originates from androstenedione. DHEAS is the major androgen produced by the adrenal gland. The adrenal gland is responsible for more than 95% of the circulating DHEAS levels. Although it is the most abundant androgen circulating in the body, it contributes minimally to serum testosterone levels. The relative strengths of androgens are listed in Table 13–1.

TABLE 13–1 Relative androgenic activity of androgens.

Steroid	Activity
Dihydrotestosterone	300
Testosterone	100
Androstenedione	10
DHEA, DHEAS	5

Reproduced with permission from Yen SSC, Jaffe RB, Barbieri RL. *Reproductive Endocrinology*. Philadelphia: Saunders/Elsevier; 1999.

The circulating estrogens include estrone, estrone sulfate, estradiol, and estriol (pregnancy). More than 95% of estradiol in the circulation is produced from the ovary. In contrast to estradiol, approximately one-half of the circulating estrone is secreted from the ovary, while the remaining is derived from peripheral conversion. The most significant precursor, androstenedione, is aromatized in the adipose tissue, hair follicles, and the liver to estrone. Estrone is also derived from estradiol through 17β-HSD type 2 activity or from estrone sulfate (steroid sulfatase). Almost all estriol is produced during pregnancy and is secreted from the placenta (see Chapter 16).

Ultimately, the serum concentration of sex steroids is dictated by the secretion rate (SR), production rate (PR), and metabolic clearance rate (MCR), as shown in Table 13–2. The SR of sex steroids from each organ determines the PR of the hormone. If the SR for a specific hormone in the ovary equals the PR, then there is no extragonadal formation. However, if there is extragonadal formation, then the PR exceeds the SR.

MCR is the volume of blood per unit of time cleared of the hormone. The MCR of sex steroids is inversely related to the affinity for sex hormone–binding globulin (SHBG) and/or albumin. Prior to excretion, steroids are conjugated to make them water soluble. The bulk of testosterone is bound to SHBG (~65%) and to a lesser extent albumin (~35%); only 1% (*free* hormone) is active and available for metabolism. The remaining androgens have negligible binding affinity for SHBG. Free testosterone may be converted to more potent androgens such as DHT or may be metabolized through androstenedione. Metabolites of androstenedione and DHT are conjugated with a sulfate or glucuronide and are excreted in the urine. The majority of estradiol is also bound, although it has less binding affinity than testosterone for SHBG (38%) and more for albumin (60%); approximately 2% is unbound. Estradiol may be directly conjugated (16α-hydroxylated or 2-hydroxylated) or is metabolized to estrone prior to conjugation. The remaining estrogens are weakly bound to proteins. Progesterone is metabolized into many intermediates prior to conjugation. Pregnanediol glucuronide is the major metabolite observed in the urine.

A variety of clinical conditions results from or cause deviations in normal secretion rates, leading to disturbances in the menstrual cycle. This chapter discusses the normal physiology and touches on several of these hormonal disturbances.

TABLE 13–2 Production rate (PR), secretion rate (SR), and metabolic clearance rate (MCR) of plasma steroids in women.

Compound	MCR of Compound in Peripheral Plasma (L/d)	Phase of Menstrual Cycle	Concentration in Plasma		PR of Circulating Compound (mg/d)	SR by Both Ovaries (mg/d)
			(nmol/L)	(µg/dL)		
Estradiol	1350	Early follicular	0.2	0.006	0.081	0.07
		Late follicular	1.2-2.6	0.033-0.070	0.445-0.945	0.4-0.8
		Midluteal	0.7	0.020	0.270	0.25
Estrone	2210	Early follicular	0.18	0.005	0.110	0.08
		Late follicular	0.5-1.1	0.015-0.030	0.331-0.662	0.25-0.50
		Midluteal	0.4	0.011	0.243	0.16
Progesterone	2200	Follicular	3.0	0.095	2.1	1.5
		Luteal	36	1.13	25.0	24.0
20α-Hydroxyprogesterone	2300	Follicular	1.5	0.05	1.1 5.8	0.8
		Luteal	7.5	0.25		3.3
17-Hydroxyprogesterone	2000	Early follicular	0.9	0.03	0.6	0-0.3
		Late follicular	6 6	0.20	4.0	3-4
		Midluteal		0.20	4.0	3-4
Androstenedione	2010		5.6	0.159	3.2	0.8-1.6
Testosterone	690		1.3	0.038	0.26	
Dehydroepiandrosterone	1640		17	0.490	8.0	0.3-3

Reproduced with permission from Greep RO, Astwood EP. *Handbook of Physiology.* Section 7, Vol. II, Part 1. American Physiological Society; 1973.

PHYSIOLOGY OF FOLLICULOGENESIS AND THE MENSTRUAL CYCLE

The menstrual cycle is regulated by complex interactions between the hypothalamic-pituitary-ovarian (HPO) axis and the uterus. Briefly, the hypothalamus secretes gonadotropin-releasing hormone (GnRH), which stimulates the pituitary to release FSH and LH. These gonadotropins then trigger the ovary to release an oocyte that is capable of fertilization. Concurrently, the ovary secretes hormones, which act on the endometrial lining of the uterus to prepare for implantation. In addition, the ovarian hormones feed back to the hypothalamus and pituitary, regulating the secretion of gonadotropins during the phases of the menstrual cycle. This complex interaction will be discussed in greater detail later. The hormonal changes associated with menstruation are summarized in Figure 13–4.

The Hypothalamic-Pituitary Axis

GnRH is the central initiator of reproduction. GnRH is a 10 amino acid peptide with a short half-life of 2 to 4 minutes. It is processed in specialized secreting neurons that originate in the olfactory placode during development and migrate to the arcuate nucleus of the medial basal hypothalamus (MBH). These neurons project from the median eminence and secrete GnRH, with inherent rhythmic behavior (pulse generator), into the portal vessels to reach the gonadotropes within the anterior pituitary. Kisspeptins (KiSS-1), which are natural ligands for the orphan G protein-coupled receptor (GPCR) known as GPR54, are key regulators in modulating GnRH activity for the onset of puberty (see Chapter 15)

as well as maintaining its activity in the adult (Figure 13–5). GnRH binds to its receptor, a member of the G protein–coupled 7-transmembrane-spanning receptor superfamily. 1,4,5–inositol triphosphate and diacylglycerol act as second-messengers for GnRH. The pulsatile frequency of GnRH secretion regulates gonadotropin synthesis and the secretion of pituitary gonadotropes (see Chapter 4).

During the late luteal-follicular phase, the slower pulsatile release of GnRH—every 90 to 120 minutes—favors FSH secretion. In response to FSH, the maturing follicle in the ovary secretes estradiol. This hormone is involved in a negative feedback loop that inhibits the release of FSH by indirectly decreasing the production of GnRH via gamma-aminobutyric acid neurons, in addition to possibly having a direct effect on the pituitary gland. Estradiol is also involved in a positive feedback loop that increases the frequency of GnRH pulses to every 60 minutes during the follicular phase and acts directly on the pituitary to stimulate LH secretion. LH stimulates the ovary to further increase estradiol production (see Two-cell Theory later). There is no further change in GnRH pulsatility at this point in the cycle; however, estradiol and other regulatory factors (see later) increase pituitary sensitivity to GnRH. This increased sensitivity results in a rapid elevation of LH production—the LH surge—which stimulates ovulation. After ovulation, the ruptured follicle (corpus luteum) secretes progesterone. This hormone is involved in a negative-feedback loop, indirectly through increased endogenous opioid activity and possibly directly through a reduction in GnRH pulsatility to every 3 to 5 hours. This favors FSH synthesis during the luteal-follicular transition. As progesterone levels fall again, GnRH pulsatility increases, favoring FSH release.
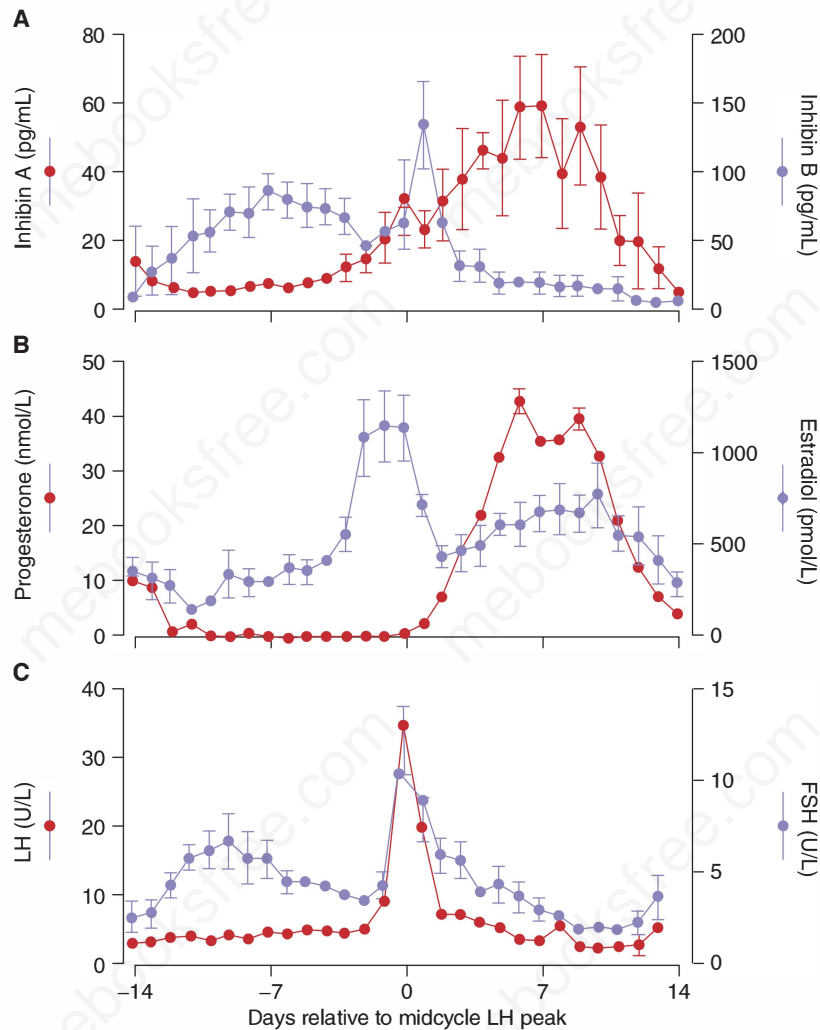
FIGURE 13–4 The endocrinology of the luteal-follicular transition in women. Data are mean ± SE of daily serum concentrations of FSH, LH, estradiol, progesterone, and immunoreactive inhibin in women with normal cycles. Note the secondary rise in plasma FSH in the late luteal phase (~2 days before menses). (Reproduced with permission from Lobo R. Menopause: Biology and Pathobiology. *Academic Press*; 2000.)

Role of the Pituitary

Gonadotropes are located in the adenohypophysis and make up approximately 10% of the cells in the pituitary. These cells synthesize and secrete FSH and LH. These hormones, thyroid-stimulating hormone (TSH) and human chorionic gonadotropin (hCG), all belong to a family of glycoprotein hormones. Gonadotropins are functional as heterodimers and are composed of an alpha and a beta subunit. The alpha subunit amino acid sequence is identical for all of the glycoprotein hormones while the beta subunits contain different amino acid sequence and confer unique specificity on the glycoproteins.

The synthesis of FSH and LH mostly occur in the same cell, yet their secretion patterns differ. The secretion of FSH is tightly linked to the expression of the FSH beta subunit. This suggests that there is minimal storage of FSH within the gonadotropes, and the majority of secretion follows more of a constitutive pathway. This is in contrast to LH secretion, in which it is first stored in organelles and then released with the appropriate trigger (regulated pathway). The different oligosaccharides on the beta subunits likely facilitate the intracellular sorting that results in different mechanisms of secretion.

The differential gene expression that leads to the production and release of gonadotropins by cells in the pituitary is influenced by GnRH and ovarian hormones through feedback loops. Slower GnRH pulse frequency enhances FSH beta subunit expression and increases LH amplitude. In turn, increased GnRH pulse frequency stimulates LH beta subunit expression while promoting FSH release. As a result, LH amplitude decreases while the mean concentration rises. Thus, ovarian steroid modification of hypothalamic GnRH pulsatility controls pituitary gonadotrophin production.

An intrapituitary network involves several factors that play a role in regulating gonadotropin synthesis and secretion. The gonadotropes produce and secrete peptides that are in
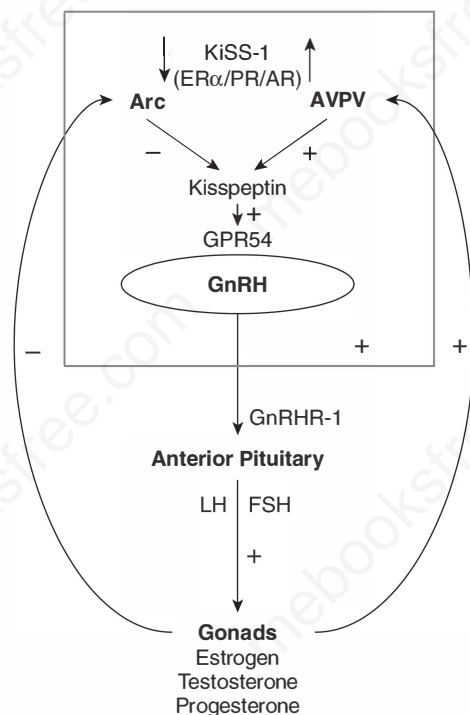
FIGURE 13–5 Kisspeptin is a key regulator of the hypothalamic-pituitary-gonadal (HPG) axis. KiSS-1 neurons in the anteroventral periventricular nucleus (AVPV) and arcuate nucleus (Arc) within the hypothalamus express kisspeptin, which has an important role in triggering the release of gonadotropin releasing hormone (GnRH). The cognate receptor for kisspeptin is GPR54. GnRH stimulates expression and release of gonadotropins (follicle stimulating hormone [FSH], luteinizing hormone [LH]), which stimulate follicular growth and steroid production in the ovaries. These steroids feedback (positive and negative) to further regulate the HPG axis and synchronize the ovarian cycle. (Reproduced with permission from Roseweir AK and Millar RP. The role of kisspeptin in the control of gonadotrophin secretion. *Hum Reprod Update.* 2009 Mar-Apr;15(2): 203-212.)

the transforming growth factor (TGF) family. Activin is a local regulatory protein that is involved in control of gonadotrope function. Slow pulses of GnRH enhance activin synthesis, which subsequently enhances FSH transcription. Follistatin, another TGF-related protein that binds to activin, is stimulated by rapid pulses of GnRH. This decreases the bioavailability of activin and consequently reduces FSH synthesis. In addition to these local modifiers, ovarian TGFs such as inhibin also modulate the expression of gonadotropins (see later).

Role of the Ovary

The ovary is intimately involved in regulating the menstrual cycle via steroid feedback to alter gonadotropin secretion. In addition, the ovary contains an intraovarian network involving factors that are synthesized locally and have a paracrine and autocrine role in the modulation of gonadotropin activity. The intraovarian regulators include the insulin-like growth factor (IGF) family, TGF

superfamily, and epidermal growth factor (EGF) family. Furthermore, it is these factors that assist in the coordination of follicular development, steroidogenesis, and ovulation.

The menstrual cycle of the ovary includes a follicular phase and a luteal phase. The follicular phase is characterized by growth of the dominant follicle and ovulation. It typically lasts 10 to 14 days. It is, however, this phase that is variable in duration and most often accounts for the variability in menstrual cycle length in ovulatory women. The luteal phase starts after ovulation and is the period when the ovary secretes hormones that are essential to accommodate conceptus implantation. This phase is relatively constant and averages 14 days (range, 12-15 days) in duration. The next section will describe the two phases in some detail.

Primordial follicles are the fundamental reproductive units that comprise the pool of oocytes that remain arrested in prophase 1 until just prior to ovulation. The meiotic arrest is maintained with the constitutive activity of a G-protein-dependent signaling pathway (Figure 13–6). Morphologically, they are composed of a primary oocyte that is surrounded by a single layer of squamous granulosa cells and a basement membrane. They have no blood supply. These primordial follicles develop between the sixth and ninth months of gestation and harbor the complete supply of ovarian follicles.

The initiation of follicular growth begins with the transition of the dormant primordial follicle into the growth phase. The exact mechanisms controlling the initial recruitment of the primordial follicle are under investigation. It is suggested that the resting follicular pool is probably under tonic inhibitory control. The initial recruitment process induces growth in some primordial follicles; other neighboring follicles remain quiescent for many months or years. It is thought that the recruitment of these follicles is a continuous process that begins once the germ pool is created and ends with follicular exhaustion. This complex process is gonadotropin-independent. Several studies have suggested that an intraovarian signaling network, involving members of the TGF superfamily, is responsible for primordial follicle recruitment. It is also known that recruitment and growth of the follicle absolutely requires cell-to-cell contact with neighboring granulosa cells and the oocyte. These cells transfer various factors, nutrients, and waste to and from the oocyte through gap junctions derived from connexins. One theory proposes that the oocyte actively influences its own fate by secreting various factors. These include two particular growth factors related to TGF-β, which are produced by the oocyte early in follicular development, growth differentiation factor (GDF)-9 and bone morphogenic protein (BMP)-15. Knockout mouse studies suggest that the oocyte induces granulosa cell proliferation through these growth factors, and the granulosa cell responds with factors (eg, follistatin, c-kit) to decrease the inhibitory influences (eg, activin A, Müllerian-inhibiting substance) and promote stimulators of oocyte growth (Figure 13–7).

Several other local factors have been described to date, and many more will be identified in the future. The ongoing search for these growth factors and hormones will ultimately elucidate the physiology of primordial follicle recruitment. There is a finite number of germ cells, and each successive recruitment further depletes the germ cell pool. Any abnormality that alters the

Pathway maintaining meiotic arrest
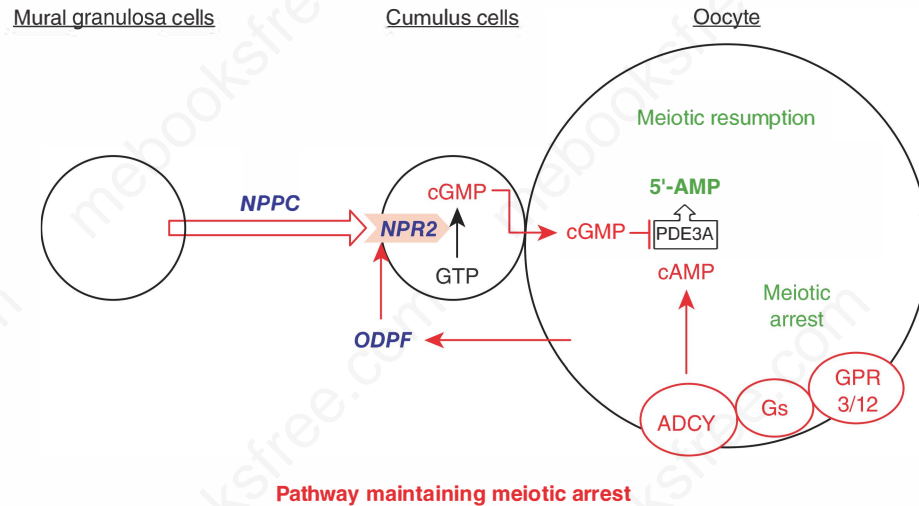
FIGURE 13–6 The meiotic arrest is maintained with the constitutive activity of a G-protein-dependent signaling pathway. Mural granulosa cells secrete natriuretic peptide, type C (NPPC). NPPC acts on its receptor (NPR2), a particulate guanylyl cyclase that is expressed on the cumulus cells, to produce cGMP. cGMP is transferred to the oocyte via gap junctional communication. In the oocyte, cGMP inhibits the degradation of another cyclic nucleotide, cyclic adenosine monophosphate (cAMP). cAMP levels are controlled primarily by the type 3 phosphodiesterase (PDE3), an enzyme whose activity is antagonized by cGMP. The oocyte is dependent on cAMP to maintain arrest. cAMP production within the oocyte is maintained through constitutive activation of G-protein-coupled receptors (ie, GPR3, GPR12) which stimulate adenylyl cyclase. (Reproduced with permission from Zhang M, Su Y-Q, Sugiura K, et al. Granulosa cell ligand NPPC and its receptor NPR2 maintain meiotic arrest in mouse oocytes. *Science*. 2010 Oct 15;330(6002):366-369.)
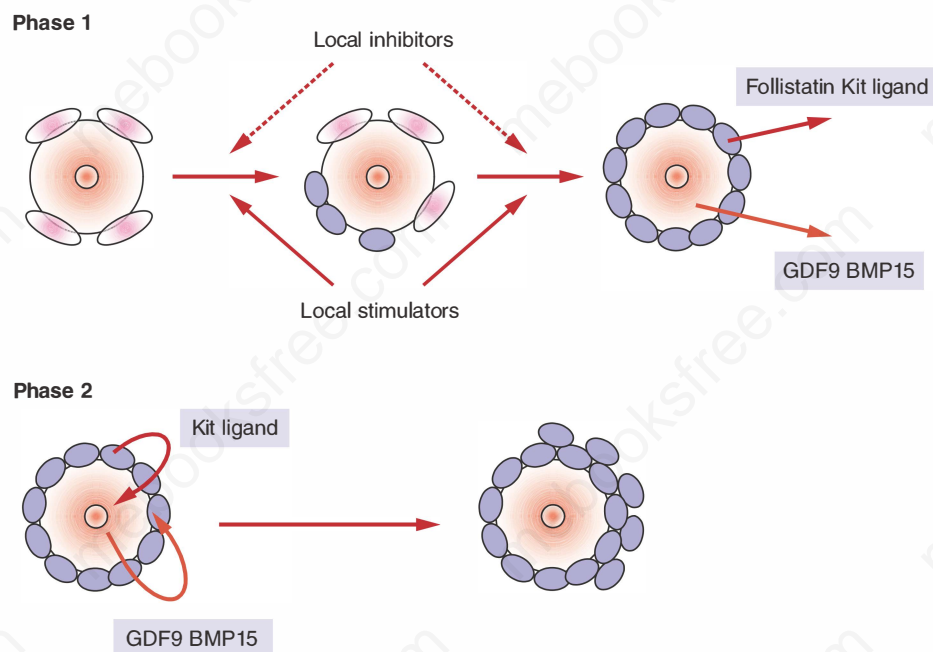
FIGURE 13–7 A theoretical model of follicular recruitment into the growth phase. Initiation of follicular growth may be regulated by many local inhibitory and stimulatory factors (ie, inhibin A, follistatin, kit ligand) and begins with transformation of the granulosa cells from squamous to cuboidal shape. Oocytes subsequently commence growth and regulate proliferation and differentiation of granulosa cells with members of the transforming growth factor β superfamily (ie, GDF-9, BMP15). Phase 1: transition of granulosa cells from flattened to cuboidal. Phase 2: oocyte growth and proliferation and differentiation of granulosa cells. (Adapted with permission from Braw-Tal R. The initiation of follicle growth: the oocyte of the somatic cells? *Mol Cell Endocrinol*. 2002 Feb 22;187(1-2):11-18.)

number of germ cells or accelerates recruitment could perhaps lead to early ovarian follicular depletion and, therefore, early reproductive failure (see section on Infertility later).

Primary follicle development is the first stage of follicular growth (Figure 13–8). Primary follicles differ from primordial follicles in several ways. The oocyte begins to grow. As growth progresses, the zona pellucida is formed. This is a thick layer of glycoprotein that is most likely synthesized by the oocyte. It completely surrounds the oocyte and forms a barrier between the oocyte and the granulosa cell layer. It serves a number of biologic functions that are critical for protection of the oocyte and conception. Finally, the granulosa cells undergo a morphologic change from squamous to cuboidal. This stage of development may last 150 days.

The progression to a **secondary follicle** includes attainment of maximal oocyte growth (120 μm in diameter), proliferation of granulosa cells, and acquisition of theca cells (see Figure 13–8). The exact mechanism involved in acquiring theca cells is not completely understood, but they are thought to be derived from the surrounding ovarian mesenchyme (stromal fibroblasts) as the developing follicle migrates to the medulla. It is the development of this layer that gives rise to the theca interna and theca externa. With theca cell development, these follicles gain an independent blood supply, although the granulosa cell layer remains avascular. In addition, the granulosa cells in the secondary follicle develop

FSH, estrogen, and androgen receptors. This phase of follicular development may take as long as 120 days, probably because of the long doubling time (>250 hours) of granulosa cells.

Further follicular development leads to the tertiary follicle or early antral phase (see Figure 13–8). This phase is characterized by the formation of an antrum or cavity in the follicle. The antral fluid contains steroids, proteins, electrolytes, proteoglycans, and an ultrafiltrate that forms from diffusion through the basal lamina. Other changes in this phase include further theca cell differentiation. Subpopulations of thecal interstitial cells develop within the theca interna, acquire LH receptors, and are capable of steroidogenesis. The granulosa cells, as well, begin to differentiate into distinct cell layers: the mural granulosa cells line the follicle wall and abut the basement membrane, and the cumulus cells are those adjacent to the oocyte. These two cell populations have different gene expression profiles and exhibit highly divergent behaviors during folliculogenesis, oocyte maturation, and ovulation.

This differentiation process is influenced by FSH and unidentified signals originating from the oocyte. It is suggested that oocyte-derived GDF-9, is an important factor in this process, where the relative concentration of GDF-9 dictates the specific subtype of granulosa cell.

An example of how the granulosa cell layers differ in activity is in the maintenance of oocyte meiotic arrest (see Figure 13–6). Both cell types acquire FSH receptors. However, the mural
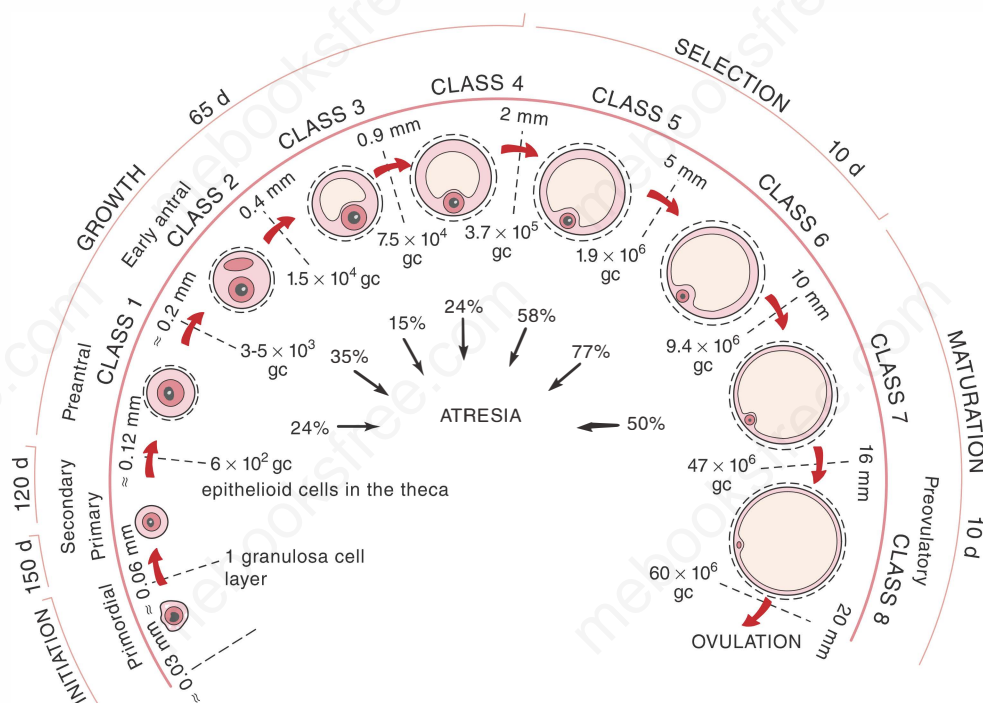
FIGURE 13–8 The chronology of folliculogenesis in the human ovary. Folliculogenesis is divided into two major periods: preantral (gonadotropin independent) and antral (FSH dependent). In the preantral period, a recruited primordial follicle develops into the primary/secondary (class 1) and early tertiary (class 2) stages, at which time cavitation or antrum formation begins. The antral period includes the small graafian (0.9-5 mm, classes 4 and 5), medium graafian (6-10 mm, class 6), large graafian (10-15 mm, class 7), and preovulatory (16-20 mm, class 8) follicles. Time required for completion of preantral and antral periods is approximately 300 and 40 days, respectively. Number of granulosa cells (gc), follicle diameter (mm), and atresia (%) are indicated. (Reproduced with permission from Gougeon A. Regulation of ovarian follicular development in primates: facts and hypotheses. *Endocr Rev.* 1996 Apr;17(2):121-155.)

granulosa cells respond to FSH by secreting an oocyte maturation inhibitor (OMI), natriuretic peptide, type C (NPPC). NPPC acts on its receptor, the type 2 natriuretic peptide receptor (NPR2), a particulate guanylyl cyclase that is expressed on the cumulus cells, to produce cGMP. cGMP is transferred to the oocyte via gap junctional communication. In the oocyte, cGMP inhibits the degradation of another cyclic nucleotide, cyclic adenosine monophosphate (cAMP). cAMP levels are controlled primarily by the type 3 phosphodiesterase (PDE3), an enzyme whose activity is antagonized by cGMP. The oocyte is dependent on cAMP to maintain arrest. cAMP production within the oocyte is maintained through constitutive activation of G-protein-coupled receptors (ie, GPR3, GPR12) which stimulate adenylyl cyclase. The oocyte maintains meiotic arrest through folliculogenesis and resumes just prior to ovulation (see later).

Both granulosa cell types—most likely in response to FSH—start producing activin, a member of the TGF family, in this phase. Activin is composed of two types of beta subunits, βA and βB, which are held together by disulfide bonds. It is the combination of these subunits that generates the various activins (activin AA [βA,βA], AB [βA,βB], or BB [βB,βB]). It is unlikely that activin has an endocrine role, because serum activin levels do not change throughout the menstrual cycle and the free-activin level in the serum is negligible. Activin's primary activity is within the ovary, where it plays an autocrine role by enhancing FSH receptor gene expression in the granulosa cells and accelerating folliculogenesis.

Follicular growth during the early antral phase occurs at a slow and constant pace. The follicle achieves a diameter of 400 μm. FSH-stimulated mitosis of granulosa cells is the major contributor to follicular growth at this stage. Until this point, follicular growth and survival are largely independent of gonadotropins. In fact, prepubertal females and women taking oral contraceptives may have follicles arrested at various phases up until this point. It is at this phase in follicular development that FSH is critical for growth and survival. If FSH does not rescue these follicles, they undergo atresia.

The morphologic follicular unit, consisting of theca cells and granulosa cells, is also a functional hormonal unit capable of substantial estrogen production (Two-cell Hypothesis; Figure 13–9). The thecal interstitial cells and granulosa cells are under the direct influence of LH and FSH, respectively. The gonadotropins increase cyclic adenosine monophosphate (cAMP) production and the activity of the transcription factor SF-1 in the respective cell types.

In the theca cell, LH increases the number of LH receptors on the cell surface and the expression and activity of StAR, P450scc, 3β-HSD-II, and P450c17 to increase androgen production. FSH increases the expression of aromatase and 17β-HSD within the granulosa cell.

The pathway of androgen synthesis is shown in Figure 13–3. Androgens may be formed in one of two ways: (1) Δ5 pathway, which entails dehydroepiandrosterone (DHEA) as the precursor to androgens, and (2) Δ4 pathway, which synthesizes androgen through 17-OH progesterone. However, in the human, the contribution of the Δ4 pathway for androgen production appears minimal. This is due to 17,20-lyase activity having a much higher affinity for 17-OH pregnenolone than 17-OH progesterone. Therefore, the main precursor to sex steroids in the human is DHEA.

The androgens—mainly androstenedione—then diffuse through the basal lamina of the follicle and are the main precursors that granulosa cells utilize for estrogen production through the activity of the aromatase enzyme. However, the pathway of estradiol biosynthesis is dictated by the type of 17β-HSD that is expressed. In humans, seven 17β-HSDs have been identified, each having different affinities for particular steroids. In the granulosa cell, it is type 1 17β-HSD that is expressed preferentially reducing estrone to yield estradiol. Type 3 17β-HSD preferentially reduces androstenedione to testosterone and is expressed in the Leydig cells and not in the ovary. However, type 5 17β-HSD is expressed in the theca cells and is likely the catalyst for the final step in testosterone synthesis from androstenedione in the ovary. In summary, the major pathway of estradiol biosynthesis within the granulosa cell entails androstenedione aromatization to estrone by aromatase, followed with a reduction to estradiol by type 1 17β-HSD.

The importance of estradiol on folliculogenesis with respect to its negative and positive feedback effects on gonadotropin secretion is well established. However, the role of estradiol in local oocyte maturation and follicular growth remains controversial. There is evidence that estrogen is synergistic with FSH during the follicular phase by increasing FSH and LH receptors, stimulating proliferation of granulosa cells and gap junctions as well as aromatase activity. The aromatase knockout (ArKO) mouse suggests a local role for estrogen. The ArKO mouse initially has large antral follicles, but after the first year, there are no remaining antral follicles or secondary follicles, and atresia is evident among the remaining primary follicles. However, the ArKO oocytes are capable of *in vitro* maturation and blastocyst formation. There is evidence that both estrogen receptors are expressed in the granulosa and theca cells. Studies with the estrogen receptor α knockout mice show these mice are infertile and lack graafian follicles. However, the estrogen receptor β knockouts are fertile but produce smaller litters. In humans, there are cases where follicular development has occurred in the absence of estradiol secretion. This was observed in a woman with CYP17α deficiency where promotion of follicular growth was noted with gonadotropins. Embryo development followed *in vitro* fertilization, but unfortunately no pregnancy resulted.

Intraovarian factors play a major modulatory role in both folliculogenesis and steroidogenesis. The oocyte-derived factor GDF-9 is expressed throughout folliculogenesis. It is thought not only to promote granulosa cell differentiation but also to have a stimulatory effect on theca cells and an inhibitory effect on luteal cell formation. The IGFs enhance the response to FSH. *In vitro* studies have shown that both IGF-1 and IGF-2 increase granulosa cell proliferation and estradiol secretion. However, it is suggested that IGF-2 rather than IGF-1 plays a dominant role in follicular maturation. This may be explained by the absence of IGF-1 expression in the granulosa cells of dominant follicles. Furthermore, women with Laron syndrome (IGF-1 deficiency) can be induced to ovulate by ovarian stimulation with gonadotropins.
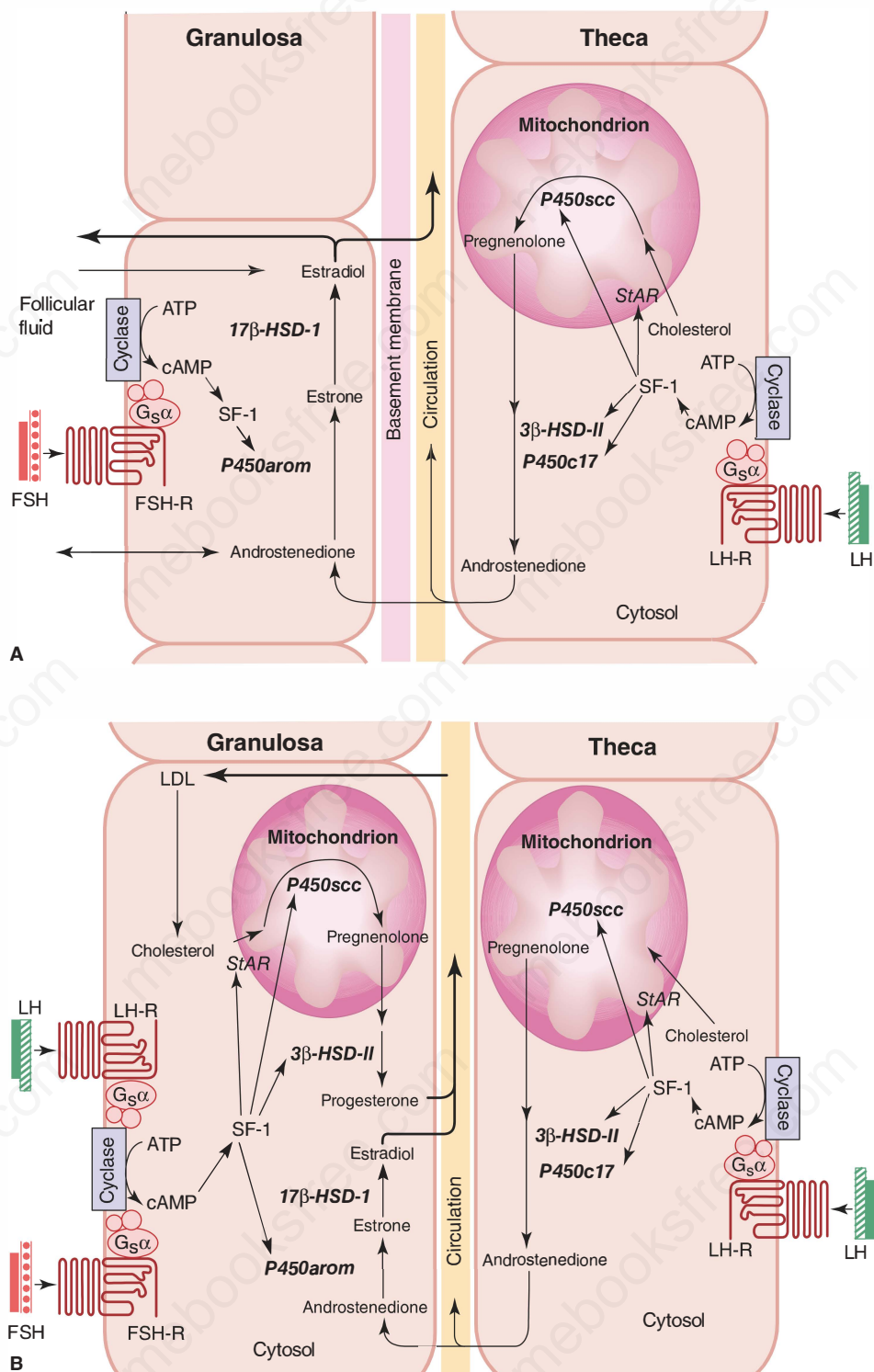
FIGURE 13–9 Two-cell hypothesis. **A.** The preovulatory follicle produces estradiol through a paracrine interaction between theca and granulosa cells. The theca cells provide the substrate, androstenedione, to the granulosa cell for estradiol production. The granulosa cells express the enzymes necessary for transformation of androstenedione to estradiol (P450 aromatase, 17β hydroxysteroid dehydrogenase-1). **B.** In the corpus luteum, granulosa-lutein cells gain vascularity, LH receptors, and the enzymes necessary for progesterone synthesis. The theca-lutein cells remain the source of androstenedione for estradiol production in granulosa-lutein cells. Gonadotropins, steroidogenic factor-1 (SF-1), and steroidogenic acute regulatory protein (StAR) play crucial roles in steroidogenesis with the ovary (ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate; FSH-R, follicle-stimulating hormone receptor; LH-R, luteinizing hormone receptor; P450 arom, P450 aromatase; 17β-HSD-I, 17β hydroxysteroid dehydrogenase-I). (Modified with permission from Larsen PR. *Williams Textbook of Endocrinology*, 10th ed. Philadelphia: Saunders; 2003.)

This scenario suggests that IGF-1 is not critical for folliculogenesis.

The granulosa cells produce other hormones (members of the TGF-β family) that regulate folliculogenesis (see Activin discussed earlier). The granulosa cells synthesize the alpha subunit, which combines with beta subunits to create a heterodimer known as inhibin A (αβA) or inhibin B (αβB). The role of inhibins in folliculogenesis and steroidogenesis is largely indirect; it suppresses pituitary FSH secretion. The serum concentrations of inhibin A and B are influenced by the menstrual rhythm. The circulating levels of inhibin A rise late in the follicular phase and remain elevated in the luteal phase whereas the concentration of inhibin B mirrors the menstrual pattern of serum FSH levels (see Figure 13–4). Although inhibin B expression is increased within the granulosa cell on stimulation with FSH, the follicular concentration of inhibin B does not change with follicular size. It has been suggested that the serum concentration of inhibin B reflects the volume of granulosa cells within the ovary, thereby serving as an indicator of the size of the growing cohort of follicles (ovarian reserve). Because inhibin B is the primary inhibitor of pituitary production of FSH in the follicular phase in the absence of estradiol, measuring baseline FSH circulating levels early in the follicular phase (cycle day 3) serves as indirect marker of ovarian reserve (see section on Infertility later).

The antral growth phase of follicular development is characterized by rapid follicle growth (1-2 mm/d) and is gonadotropin-dependent. In response to FSH, the antral follicle rapidly grows to a diameter of 20 mm, primarily as a result of accumulation of antral fluid. The theca interna continues to differentiate into interstitial cells that produce increasing amounts of androstenedione for aromatization to estrone. The granulosa cell layers have continued to differentiate from each other. The mural layer, through the action of FSH, acquires LH receptors. This differs from the cumulus layer, which has low or undetectable levels of LH receptors. The final progression to a mature graafian follicle is a selection process that in most cases generates one dominant follicle destined for ovulation.

The selection process actually begins in the midluteal phase of the previous cycle. The rise in estrogen level that is generated by the preovulatory follicle augments FSH activity within the follicle, while exerting negative feedback on the pituitary release of FSH. The decrease in pituitary release of FSH results in withdrawal of gonadotropin support from the smaller antral follicles, promoting their atresia. The dominant follicle continues to grow, despite decreasing levels of FSH, by accumulating a greater mass of granulosa cells with more FSH receptors. Increased vascularity of the theca cells allows preferential FSH delivery to the dominant follicle despite waning FSH levels. Increased estrogen levels in the follicle facilitate FSH induction of LH receptors on the granulosa cells, allowing the follicle to respond to the ovulatory surge of LH levels. Without estrogen, LH receptors do not develop on the granulosa cells.

The generation of the LH surge is absolutely required for oocyte maturation (meiotic resumption). The amplified production of LH in the midcycle is due to an increase in pituitary sensitivity to GnRH. The sensitization is mediated by the positive feedback effect of the exponential rise in estrogen and possibly inhibin A. This surge results in the resumption of meiosis I in the oocyte with release of a polar body just prior to ovulation. Evidence suggests the mechanism of meiotic resumption commences with the pre-ovulatory LH surge which decreases C-type natriuretic peptide secretion and consequently leads to a decrease in cGMP production within the granulosa cells. As a result, cGMP concentrations are decreased within the oocyte, and oocyte-specific PDE3 activity increases which decreases cAMP concentrations. The decrease in cAMP leads to activation of an enzyme cascade that allows a burst in Ca^{++} and meiosis resumes.

Just prior to the ovulation, progesterone production increases, and this may be responsible, at least in part, for the midcycle peak in FSH and the coordination of the LH surge. The FSH peak stimulates the production of an adequate number of LH receptors on the mural granulosa cells for luteinization. FSH, LH, and progesterone induce expression of proteolytic enzymes that degrade the collagen in the follicular wall, thereby making it prone to rupture. Prostaglandin (PG) production increases and may be responsible for contraction of smooth muscle cells in the ovary, aiding the extrusion of the oocyte.

The LH surge lasts approximately 48 to 50 hours. Thirty-six hours after onset of the LH surge, ovulation occurs. The feedback signal to terminate the LH surge is not known. Perhaps the rise in progesterone production results in a negative feedback loop and inhibits pituitary LH secretion by decreasing the pulsatility of GnRH. In addition, just prior to ovulation, LH downregulates its own receptors, which decreases the activity of the functional hormonal unit. As a result, estradiol production decreases.

Following ovulation and in response to LH, the mural granulosa cells and thecal interstitial cells that remain in the ovulated follicle differentiate into granulosa lutein and theca lutein cells, respectively, to form the corpus luteum. In addition, LH induces the granulosa lutein cells to produce vascular endothelial growth factor (VEGF), which plays an important role in developing the corpus luteum vascularization. This neovascularization penetrates the basement membrane and provides the granulosa lutein cells with LDL for progesterone biosynthesis. After ovulation, the luteal cells upregulate their LH receptors by an unknown mechanism. This is critical in that it allows basal levels of LH to maintain the corpus luteum. Rescue of the corpus luteum with hCG from the developing conceptus works through the LH receptor, which is vital for embryonic life.

Hormone production by the corpus luteum requires the cooperation of theca lutein and granulosa lutein cells, much like the preovulatory follicle (see Figure 13–9). In response to LH and hCG, the theca lutein cells increase the expression of all the enzymes responsible for androstenedione production (see earlier). Aromatase activity is increased in the granulosa lutein cells by LH to aromatize the androgens to estrogen. A notable difference in the granulosa lutein cells, as opposed to the preovulatory granulosa cells, is that LH also induces the expression of P450scc and 3β-HSD, which enables the cell to synthesize progesterone. The secretion of progesterone and estradiol is episodic and correlates with the LH pulses. FSH has minimal influence on progesterone production but continues to stimulate estrogen production during

the luteal phase. The progesterone levels rise and reach a peak on approximately day 8 of the luteal phase. The luteal phase lasts approximately 14 days.

The corpus luteum starts to undergo luteolysis (programmed cell death) approximately 9 days after ovulation. The mechanism of luteal regression is not completely known. Once luteolysis begins, there is a rapid decline in progesterone levels. A number of studies suggest that estrogen has a role in luteolysis. It has been shown that direct injection of estrogen into the ovary containing a corpus luteum induces luteolysis and a fall in progesterone levels. Experimental data suggest that there is increased aromatase activity in the corpus luteum just prior to luteolysis. The rise in aromatase activity is secondary to gonadotropin (FSH and LH) stimulation, although later in the luteal phase, FSH probably plays a more important role. Consequently, estrogen production increases, and this decreases 3β-HSD activity. This may result in a decline in progesterone levels and lead to luteolysis. Furthermore, local modifiers such as oxytocin, which is secreted by luteal cells, have been shown to modulate progesterone synthesis. Other evidence supports the role of PGs in luteolysis. Experimental data suggest that $PGF_{2\alpha}$, which is secreted from the uterus or ovary during the luteal phase, stimulates the synthesis of cytokines such as tumor necrosis factor; this causes apoptosis and, therefore, may be linked to corpus luteum degeneration.

The process of luteolysis is known to involve proteolytic enzymes. Evidence suggests that matrix metalloproteinase (MMP) activity is increased during luteolysis. hCG is a known modulator of MMP activity. This may play an important role in early pregnancy, when hCG rescues the corpus luteum and prevents luteal regression. However, in the absence of pregnancy, the corpus luteum regresses, resulting in a decrease in progesterone, estradiol, and inhibin A levels. The decrease in these hormones allows for increased GnRH pulsatility and FSH secretion. The rise in FSH will rescue another cohort of follicles and initiate the next menstrual cycle.

Role of the Uterus

The sole function of the uterus is to accommodate and support a fetus. Furthermore, it is the endometrium, the lining of the uterine cavity, that differentiates during the menstrual cycle so that it can support and nourish the conceptus. Histologically, the endometrium is made up of an epithelium, composed of glands and a stroma that contains stromal fibroblasts and extracellular matrix. The endometrium is divided into two layers based on morphology: the basalis layer and the functionalis layer. The basalis layer lies adjacent to the myometrium and contains glands and supporting vasculature. It provides the components necessary to develop the functionalis layer. The functionalis is the dynamic layer that is regenerated every cycle. More specifically, it is this layer that can accommodate implantation of the blastocyst.

During the menstrual cycle, the endometrium responds to hormones secreted from the ovaries. Somewhat like the other endocrine organs, it contains a network of local factors that modulate hormonal activity. The endometrial phases are coordinated with ovulatory phases. During the follicular phase, the endometrium goes through the proliferative phase. It begins with the onset of menses and ends at ovulation. During the luteal phase, the endometrium undergoes the secretory phase. It starts at ovulation and ends just before menses. If implantation does not occur, a degenerative phase follows the secretory phase within the endometrium. It is this phase that results in menstruation. The next section will discuss the phases of the endometrium in more detail.

During the follicular phase, the ovary secretes estrogen, which stimulates the glands in the basalis to initiate formation of the functionalis layer. Estrogen promotes growth by enhancing gene expression of cytokines and a variety of growth factors, including EGF, TGFα, and IGF. These factors provide a microenvironment within the endometrium to modulate the effects of hormones. At the beginning of the menstrual cycle, the endometrium is thin—usually less than 2 mm in total thickness. The endometrial glands are straight and narrow and extend from the basalis toward the surface of the endometrial cavity. As the epithelium and the underlying stroma develop, they acquire estrogen and progesterone receptors. The spiral blood vessels from the basalis layer extend through the stroma to maintain blood supply to the epithelium. Ultimately, the lining (functionalis) surrounds the entire uterine cavity and achieves a thickness of 3 to 5 mm in height (total thickness 6-10 mm). This phase is known as the proliferative phase.

After ovulation, the ovary secretes progesterone, which inhibits further endometrial proliferation. This mechanism may be mediated by antagonizing estrogen effects. Progesterone downregulates estrogen receptors in the epithelium and regulates estradiol metabolism within the endometrium by stimulating 17β-HSD activity and converting estradiol into the weaker estrogen, estrone. During the luteal phase, the glandular epithelium accumulates glycogen and begins to secrete glycopeptides and proteins—along with a transudate from plasma—into the endometrial cavity. It is this fluid that provides nourishment to the free-floating blastocyst. Progesterone also stimulates differentiation of the endometrium and causes characteristic histologic changes. The glands become progressively more tortuous, and the spiral vessels coil and acquire a corkscrew appearance. The underlying stroma becomes very edematous as a result of increased capillary permeability, and the cells begin to appear large and polyhedral, with each cell developing an independent basement membrane. This process is termed predecidualization. These cells are very active and respond to hormonal signals. They produce PGs along with other factors that play an important role in menstruation, implantation, and pregnancy. This phase is known as the secretory phase.

If there is no implantation of an embryo, the endometrium enters the degenerative phase. Estrogen and progesterone withdrawal promotes PG production—$PGF_{2\alpha}$ and PGE_2. These PGs stimulate progressive vasoconstriction and relaxation of the spiral vessels. These vasomotor reactions lead to endometrial ischemia and reperfusion injury. Eventually, there is hemorrhage within the endometrium with subsequent hematoma formation. The progesterone withdrawal triggers MMP activity, which facilitates degradation of the extracellular matrix. As ischemia and degradation progress, the functionalis becomes necrotic and sloughs away as

menstruum consisting of endometrial tissue and blood. The amount of blood lost in normal menses ranges from 25 mL to 60 mL. Although $PGF_{2\alpha}$ is a potent stimulus for myometrial contractility and limits postpartum bleeding, it has minimal impact on cessation of menstrual bleeding. The major mechanisms responsible for limiting blood loss involve the formation of thrombin-platelet plugs and estrogen-induced healing of the basalis layer by reepithelialization of the endometrium, which begins in the early follicular phase of the next menstrual cycle.

If conception takes place, implantation can occur in the endometrium during the midsecretory (midluteal) phase, at which time it is of sufficient thickness and full of sustenance. The syncytiotrophoblast subsequently secretes hCG, which rescues the corpus luteum and maintains progesterone secretion, essential for complete endometrial decidual development.

In summary, the ovary has two phases during the menstrual cycle: the follicular phase and the luteal phase. The endometrium has three phases and is synchronized by the ovary. The complex feedback loops between the ovary and the hypothalamic-pituitary axis regulate the menstrual cycle. During the follicular phase, the ovary secretes estradiol, which stimulates the endometrium to undergo the proliferative phase. After ovulation (luteal phase), the ovary secretes estrogen and progesterone, which maintains the endometrial lining and promotes the secretory phase. In a nonpregnant cycle, luteolysis occurs, resulting in cessation of hormone production. This hormone withdrawal results in the degenerative phase and the onset of menses.

MENSTRUAL DISTURBANCES

AMENORRHEA

The mean age of menarche is 12.8 years. Normally, this event marks the completion of puberty. The onset of regular cyclicity in the menstrual cycle is determined by the duration of the maturation process of the HPO axis, which is quite variable. As a result, anovulation can occur in 50% to 80% of girls 2 years after menarche, and in more than 20% of girls, it can persist until 5 years after menarche. This period of time can be frustrating for many girls and their parents. The evaluation should include a clinical examination and reassurance. If positive findings are noted, limited, pertinent diagnostic tests (see later) are indicated.

Amenorrhea can be defined as either the absence of menarche by age 16 or no menses for more than three cycles in an individual who has previously had cyclic menses. The definition, although arbitrary, nonetheless gives a general guideline to the clinician for further evaluation. Although amenorrhea does not cause harm, in the absence of pregnancy, it may be a sign of genetic, endocrine, and/or anatomic abnormalities. If the outflow tract is intact, amenorrhea is most likely the result of disruption in the HPO axis. These aberrations can affect any level of control in the menstrual cycle and, thus, result in menstrual abnormalities.

Amenorrhea was formerly classified as primary or secondary depending on whether the individual had experienced menses in the past. This classification may lead to misdiagnosis of the cause of amenorrhea. Although primary amenorrhea is more often associated with genetic and anatomic abnormalities, each individual should be assessed by means of the history and clinical findings, including the presence or absence of secondary sexual characteristics (Table 13–3). The causes of amenorrhea are grouped according to the level of involvement in the regulatory systems that govern normal menstrual activity (ie, hypothalamic, pituitary, ovarian, and uterine). An algorithm for the workup of amenorrhea or oligomenorrhea in the presence of secondary sexual characteristics is illustrated in Figure 13–10.

HYPOTHALAMIC AMENORRHEA

Isolated GnRH Deficiency

The hypothalamus is the source of GnRH, which directs the synthesis and secretion of pituitary gonadotropins. Dysfunction at this level leads to hypogonadotropic hypogonadism or eugonadotropic hypogonadism. Disorders of GnRH production can be due to genetic or hormonal causes within the hypothalamic-pituitary-adrenal axis (Figure 13–11) and result in a wide range of clinical manifestations. The individual's appearance is dependent on the age at onset and the degree of dysfunction.

Isolated GnRH deficiency results in hypogonadotropic hypogonadism. Female patients present with amenorrhea, and females and males present with absent or incomplete pubertal development secondary to absent or diminished sex steroids (estradiol in females, testosterone in males). They have normal stature with a eunuchoidal body habitus. Because the adrenal glands are unaffected by the absence of GnRH, body hair distribution is not affected.

A. Genetic origin Several genetic lesions associated with GnRH deficiency have been described. The best characterized form of GnRH deficiency is Kallmann syndrome, which involves the *Kal-1* gene as well as a number of other genes. The *Kal-1* gene normally codes for anosmin, an adhesion molecule that appears to be involved in the migration of GnRH and olfactory neurons from the olfactory placode to the hypothalamus. The *Kal-1* gene is located on the short arm of the X chromosome. Most cases of Kallmann syndrome are sporadic, although the disorder has also been observed to have a familial pattern, and most often it is transmitted by X-linked recessive inheritance. Autosomal recessive and dominant patterns have been reported but are much less common. When mutations exist in the *Kal-1* gene, there may be associated defects, including anosmia and, less frequently, midline facial defects, renal anomalies, and neurologic deficiency. The disorder affects both sexes, but because of the X-linked inheritance pattern it is more common in boys. Unlike males, the specific genetic mutations in the *Kal-1* gene in females with hypogonadotropic hypogonadism have not been identified, suggesting there are other genetic mutations that cause this disorder. Recently, it was determined that heterozygous mutations in fibroblast growth factor receptor 1 (*FGFR1*) gene, known as the *Kal-2*, were identified in cases of Kallmann syndrome. *FGFR1* and *Kal-1* are coexpressed at the same sites during development and are thought to

TABLE 13–3 Assessment of patients with amenorrhea.

I. **Absent breast development; uterus present**
   A. Gonadal failure
      1. Gonadal agenesis
      2. Gonadal dysgenesis
         a. 45,X (Turner syndrome)
         b. 46,X abnormal X (eg, short- or long-arm deletion)
         c. Mosaicism (eg, X/XX, X/XX/XXX)
         d. 46,XX or 46,XY (Swyer syndrome) gonadal dysgenesis
   B. Defects in estrogen biosynthesis (46,XX)
      1. 17,20-Lyase deficiency
      2. CYP17α deficiency
   C. Hypothalamic failure secondary to inadequate GnRH release
      1. Insufficient GnRH secretion
         a. FHA
         b. Anorexia nervosa and bulimia
         c. CNS neoplasm (craniopharyngioma, gliomas)
         d. Excessive exercise
         e. Constitutional delay
      2. Inadequate GnRH synthesis (Kallmann syndrome)
      3. Developmental anatomic abnormalities in central nervous system
   D. Pituitary failure
      1. Isolated gonadotropin insufficiency
      2. GnRH resistance
      3. Pituitary tumors (hyperprolactinemia)
      4. Pituitary insufficiency
         a. Infections (mumps, encephalitis)
         b. Newborn kernicterus
      5. Prepubertal hypothyroidism
II. **Breast development; uterus absent**
   A. Androgen resistance (androgen insensitivity syndrome)
   B. Congenital absence of uterus (uterovaginal agenesis)
III. **Absent breast development; uterus absent**
   A. Defects in testosterone biosynthesis (46,XY)
      1. 17,20-Lyase deficiency
      2. CYP17α deficiency
      3. 17β-Hydroxysteroid dehydrogenase deficiency
   B. Testicular regression syndrome (46,XY)

IV. **Breast development; uterus present**
   A. Pregnancy
   B. Hypothalamic etiology
      1. FHA
      2. Anorexia nervosa and bulimia
      3. Psychogenic (depression)
      4. CNS neoplasm
      5. Chronic disease
   C. Pituitary etiology
      1. Pituitary tumors (hyperprolactinemia)
      2. Pituitary insufficiency
         a. Hypotensive event (Sheehan syndrome)
         b. Infections
         c. Autoimmune destruction
         d. Iatrogenic (surgery, radiation)
   D. Ovarian etiology
      1. POF
         a. Mosaicism (46,XX/XO,XX/XY)
         b. Autoimmune destruction
         c. Iatrogenic (radiation, chemotherapy)
         d. Fragile X syndrome
         e. Infections
      2. Resistant ovarian syndrome (Savage syndrome)
   E. Chronic estrogenized anovulation
      1. Hyperandrogenic
         a. PCOS
         b. Nonclassical congenital adrenal hyperplasia
         c. Cushing syndrome
         d. Androgen secreting tumors
      2. Other
         a. Adrenal insufficiency
         b. Thyroid disorders
   F. Outflow tract
      1. Congenital abnormalities
         a. Transvaginal septum
         b. Imperforate hymen
      2. Asherman syndrome

have a functional interaction. Mutations in the *Kiss-1* derived peptide receptor, GPR54, have also been identified in these patients. Several mechanisms of how GPR54 may regulate gonadotropin production via GnRH function or production have been proposed. Several studies have shown that females with presumed Kallmann syndrome demonstrate variable responses to exogenous GnRH administration, which suggests a GnRH receptor defect. In fact, mutations in the GnRH receptor have been identified in both sexes and are inherited in an autosomal recessive fashion.

Management of hypogonadotropic hypogonadism involves scheduled hormone replacement therapy (HT) to stimulate the development of secondary sexual characteristics and increase bone mineral density (BMD). If pregnancy is desired, treatment involves the administration of pulsatile GnRH or gonadotropin treatment.

B. Endocrine causes

1. **Functional hypothalamic amenorrhea**—Functional hypothalamic amenorrhea is one of the most common types of amenorrhea and accounts for 15% to 35% of cases. It is an endocrine disorder, although the exact mechanism has not been definitively determined. It is characterized by a reduced GnRH drive (decrease in pulse frequency and amplitude),

leading to low or low-normal serum levels of FSH and LH and resulting in anovulation. The ratio of serum FSH to LH in these patients is often equivalent to that of a prepubertal female with a relative FSH dominance.

The adipocyte hormone, leptin, has been implicated in the development of this disorder. Leptin is an important nutritional satiety factor, but it is also necessary for maturation of the reproductive system. The potential link to the reproductive system is thought to be through leptin receptors, which have been identified in the hypothalamus and gonadotropes. This is supported by the observation that leptin can stimulate GnRH pulsatility and gonadotropin secretion. Several studies suggest that women with functional hypothalamic amenorrhea have lower serum leptin levels in comparison with eumenorrheic controls. This relative deficiency may lead to a dysfunctional release of GnRH, possibly mediated through a kisspeptin pathway, and subsequent development of functional hypothalamic amenorrhea.

Abnormal, and often subtle, activation of the hypothalamic-pituitary-adrenal axis is associated with functional hypothalamic amenorrhea. The inciting event may be excessive production of corticotropin-releasing hormone (CRH), which has been shown to decrease the pulse frequency of GnRH and increase cortisol levels *in vivo*. In contrast, another study
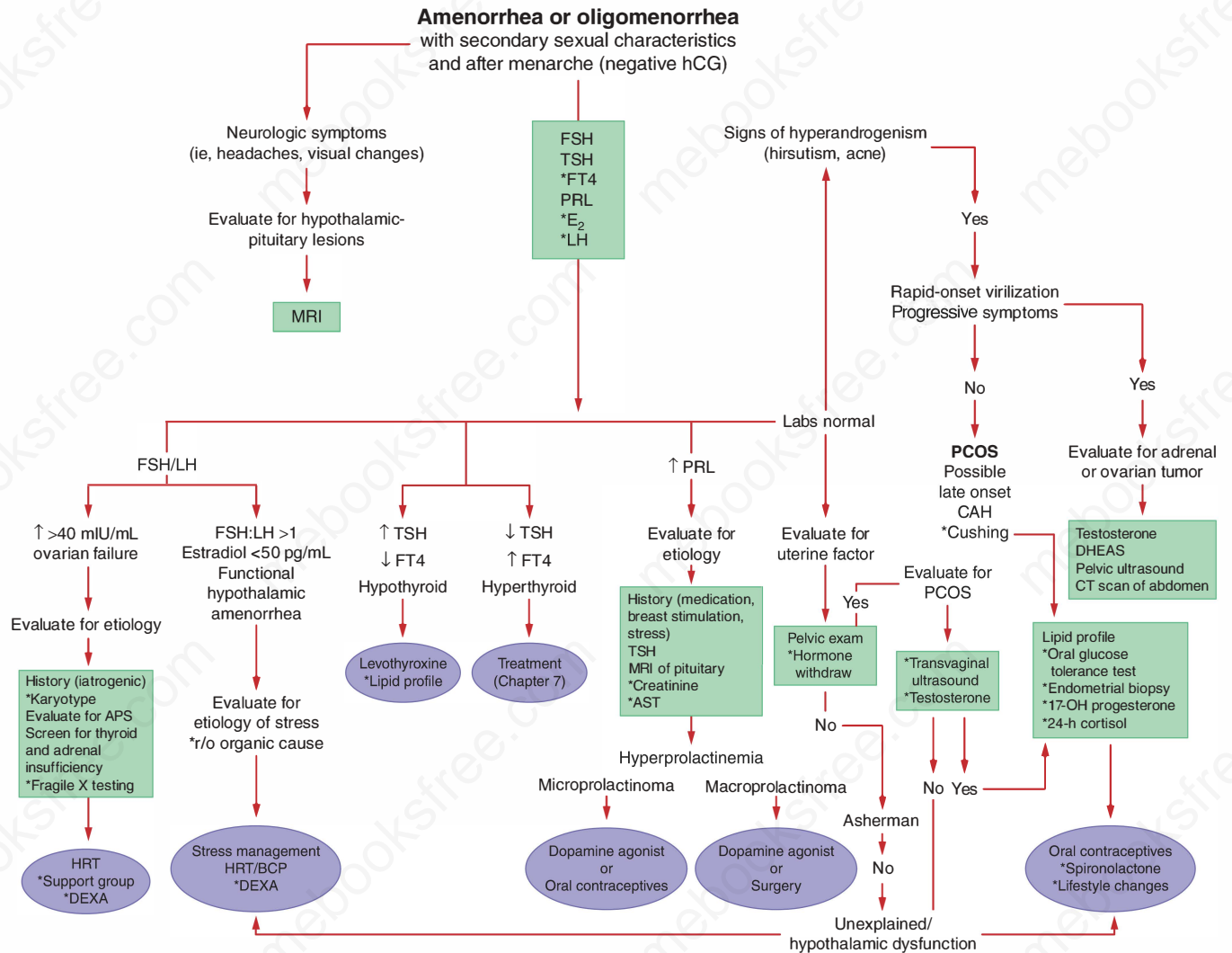
FIGURE 13–10 An algorithm for women who have developed secondary sexual characteristics, have experienced menarche, and are presenting with amenorrhea or oligomenorrhea. A negative human chorionic gonadotropin test documents the nonpregnant state. All women should be screened for known causes of oligomenorrhea, whether or not hyperandrogenism is present. Hyperandrogenism suggests either polycystic ovarian syndrome (PCOS) or late-onset congenital adrenal hyperplasia (CAH). If rapid in onset, progressive virilization requires an evaluation for an androgen-secreting tumor. Management of late-onset CAH and PCOS is similar. Clinical history and physical findings suggesting hypoestrogenism, stress, or thyroid dysfunction are indications for measurement of estradiol (E_2), luteinizing hormone (LH), and free thyroid hormone (FT4) levels in plasma (APS, autoimmune polyglandular syndrome). *As indicated (refer to text).

suggests that although acute elevations of CRH can suppress GnRH release, this suppression cannot be maintained with CRH alone. The cause of functional hypothalamic amenorrhea often remains unclear, but the associated hypercortisolemia suggests that it is preceded by psychologic stress, strenuous exercise, or poor nutrition. There is support for the concept that these factors may act synergistically to further suppress GnRH drive. In fact, patients with functional hypothalamic amenorrhea, resulting from psychologic stress, are usually high achievers who have dysfunctional coping mechanisms when dealing with daily stress. The severity of hypothalamic suppression is reflected by the clinical manifestations. The significant interpatient variability in the degree of psychologic or metabolic stress required to induce a menstrual disturbance explains the heterogeneity of clinical presentations, ranging from luteal phase defects to anovulation with erratic bleeding to amenorrhea.

Functional hypothalamic amenorrhea is reversible. Interestingly, the factors that have predicted the rate of recovery are body mass index and basal cortisol levels. When patients recover, ovulation is preceded by return of cortisol levels to baseline. Some experts have shown that cognitive behavioral therapy, teaching the patient how to cope with stress—and nutritional consultation—reverse this condition. Complete reversal may be less likely if the functional insult occurs during the period of peripubertal maturation of the HPO axis.

The diagnosis of functional hypothalamic amenorrhea can be made if the FSH-LH ratio is greater than 1 in the presence of hypoestrogenemia. However, a minor disturbance of hypothalamic dysfunction may present with normal laboratory
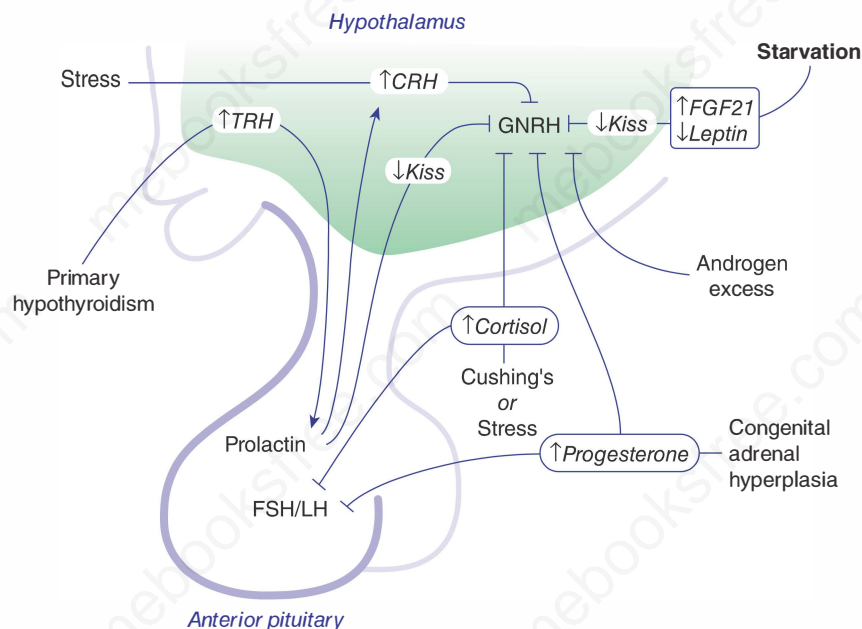
FIGURE 13–11 Potential hormonal mediators of amenorrhea. There are several conditions that influence the production of hormones which modulate GnRH release. Conditions (ie, stress, Cushings disease) which increase CRH and/or cortisol levels suppress gonadotropin-releasing hormone (GnRH) release and/or decrease LH responsiveness to GnRH. Low leptin levels and likely increased FGF-21 levels (ie, starvation) possibly suppress GnRH release by decreasing kisspeptin expression. Increased thyrotropin releasing hormone (TRH) levels (ie, primary hypothyroidism) stimulate prolactin release. Hyperprolactinemia suppresses GnRH release by decreasing kisspeptin expression, and possibly through a CRH-mediated pathway. High androgen and progesterone levels may be seen in congenital adrenal hyperplasia (CAH), and these hormones likely suppress GnRH release. (Reproduced with permission from Fourman LT, Fazeli PK. Neuroendocrine causes of amenorrhea—an update. *J Clin Endocrinol Metab*. 2015 Mar;100(3):812-824.)

findings, a clinical history that coincides with stress, and a negative evaluation for other causes of anovulation. Interestingly, most of these patients, despite hypoestrogenism, do not have symptoms. The estrogen status should still be evaluated given the strong correlation between hypoestrogenemia and the development of osteoporosis. Estrogen status can be determined by means of the progesterone withdrawal test or by measurement of serum estradiol (<50 pg/mL). If there is no withdrawal or the estradiol is less than 50 pg/mL, HRT with combination contraceptive hormones (or traditional HRT) should be instituted. If withdrawal bleeding occurs, any cyclic progestin-containing therapy is adequate to combat unopposed estrogen and the development of endometrial hyperplasia.

2. **Amenorrhea in the female athlete**—Hypothalamic dysfunction has been observed in female athletes. Competitors in events such as gymnastics, ballet, marathon running, and diving can show menstrual irregularities ranging from luteal phase defects to amenorrhea. The **female athletic triad** as defined by the American College of Sports Medicine is characterized by disordered eating, amenorrhea, and osteoporosis. The associated nutritional deficiencies can lead to impaired growth and delayed sexual maturation. The neuroendocrine abnormalities are similar to those of women with functional hypothalamic amenorrhea.

These patients have very low body fat, often below the tenth percentile. There is evidence that a negative correlation between body fat and menstrual irregularities exists. In addition, there appears to be a critical body fat level that must be present in order to have a functioning reproductive system. Several studies have shown that these amenorrheic athletes have significantly lower serum leptin levels, which further supports leptin's role as a mediator between nutritional status and the reproductive system. The strenuous exercise these athletes engage in amplifies the effects of the associated nutritional deficiency. This synergism causes severe suppression of GnRH, leading to the low estradiol levels.

Amenorrhea alone is not harmful. However, low serum estradiol over a period of time may lead to osteoporosis and delayed puberty. An analysis of estrogen status may be obtained with measurement of serum estradiol levels or with the progestin withdrawal test (see earlier). If estrogen is low, BMD should be assessed by dual-energy x-ray absorptiometry (DXA) scan. All patients diagnosed with female athletic triad need combination contraceptive therapy or HT.

3. **Amenorrhea associated with eating disorders**—Anorexia nervosa is a disorder characterized by relentless dieting in pursuit of a thin body habitus. Approximately 95% of cases occur in females, and the onset is chiefly in adolescence. The clinical features include extreme weight loss leading to a body weight less than 85% of normal for age and height, a distorted body image, and intense fear of gaining weight. These patients usually have a preoccupation with food and are hyperactive, with an obsessive-compulsive personality. The associated symptoms include hypothermia, mild bradycardia, dry skin, constipation, and symptoms of hypoestrogenemia. Furthermore, as part of

the diagnostic criteria, they must experience at least 3 months of no menses.

The dysfunction in the neuroendocrine system is similar to but often more severe than that described in association with functional hypothalamic amenorrhea. The severe reduction in GnRH pulsatility leads to suppression of FSH and LH secretion, possibly to undetectable levels, and results in anovulation and low serum estradiol levels. Given the severe psychologic and metabolic stress experienced by these individuals, the hypothalamic-pituitary-adrenal axis is activated. The circadian rhythm of adrenal secretion is maintained, but both cortisol production and plasma cortisol levels are persistently elevated secondary to increased pituitary secretion of adrenocorticotropic hormone (ACTH). Serum leptin levels in these individuals are significantly lower than normal healthy controls and correlate with percentage of body fat and body weight. A rise in leptin levels in response to dietary treatment is associated with a subsequent rise in gonadotropin levels. This further suggests leptin's role as a potential link between energy stores and the reproductive system.

The self-induced starvation state associated with anorexia nervosa leads to additional endocrine abnormalities not observed in other causes of hypothalamic amenorrhea. For instance, thyroid hormone metabolism is altered. TSH and T_4 levels are in the low-normal range, but T_3 levels are usually below normal. This is attributable to decreased peripheral conversion of T_4 to T_3 and increased conversion of T_4 to the metabolically inactive thyroid hormone, reverse T_3—a change that often resembles other states of starvation. This may be a protective mechanism in that the relative hypothyroid state attempts to reduce basal metabolic function in response to a highly catabolic state.

Bulimia occurs in about half of patients with anorexia and is defined as binge eating followed by self-induced purging. Not all bulimics have low body weight—in fact, normal-weight bulimic individuals are much more common. These patients also have a variety of neuroendocrine aberrations—often to a lesser degree than those with anorexia—which also lead to menstrual disturbances. Leptin levels are lower than in matched controls but not as low as in individuals with anorexia nervosa. They also have neurotransmitter abnormalities—notably low serotonin levels—which might help explain the often coexisting psychologic difficulties.

Anorexia nervosa is a life-threatening illness with a significant mortality rate due to its metabolic consequences. Anorexic patients should be considered for inpatient therapy and management with a multidisciplinary approach that includes nutritional counseling and psychotherapy. Force-feeding may be necessary in some patients. If weight gain cannot be achieved with oral intake, meals may need to be supplemented by enteral or parenteral feeding. Because anorexia nervosa is a hypoestrogenic state and there is a high potential for the development of osteoporosis, all patients should receive hormone therapy either in the form of HT or combination contraceptive pills. It is important to note that many patients with anorexia will continue to have reduced bone size regardless of sex steroid administration presumably due to the concurrent malnutrition and metabolic compromise. Weight rehabilitation and recovery is associated with bone accretion; however, the loss of bone may not be completely reversible.

In summary, the hypothalamic amenorrhea endocrine syndromes are probably a continuum of disordered eating and nutritional deficiencies resulting in increasingly severe abnormalities in the reproductive system. Furthermore, the age at onset affects the potential complications of these disorders. If low estradiol levels are present before age 20, bone mineralization may be profoundly affected, because this period is critical for building peak bone mass. In addition, if these conditions occur prior to puberty, it may result in stunted growth and delayed development of secondary sexual characteristics.

C. Anatomic causes   Numerous anatomic abnormalities within the central nervous system can result in menstrual disturbances. These include developmental defects, brain tumors, and infiltrative disorders. The most common anatomic lesion associated with delayed puberty and amenorrhea is a craniopharyngioma. It is derived from Rathke pouch and extends into the hypothalamus, pituitary, and third ventricle. The symptoms include headaches, visual loss, and hypoestrogenism.

Infiltrative disorders that involve the hypothalamus are uncommon but can result from systemic diseases, including sarcoidosis, histiocytosis, hemochromatosis, and lymphoma. These diseases do not initially present with amenorrhea. However, in the presence of these diseases, the hypothalamus may be affected, so they should be part of the differential diagnosis of amenorrhea.

PITUITARY AMENORRHEA

There are a few genetic mutations affecting the pituitary that cause amenorrhea. Rare autosomal recessive mutations may cause deficiencies in FSH, LH, TSH, prolactin (PRL), and growth hormone (GH). The clinical manifestations may include delayed puberty, a hypoestrogenic state, and infertility.

A. Genetic causes   A deficiency in FSH and LH may be a result of *GnRH receptor* gene mutations. Such mutations are primarily compound heterozygous mutations that affect GnRH receptor–dependent signal transduction. The phenotype of these individuals is similar to that of those with isolated GnRH deficiency. In fact, some investigators speculate that these receptor mutations may be the cause of isolated GnRH deficiency in women, given that no mutations have yet been identified in the ligand or *Kal-1* gene. The estimated prevalence of GnRH receptor mutations in women with hypothalamic amenorrhea is 2%. In a family with other affected females, the prevalence is 7%.

Other rare genetic defects have been associated with amenorrheic women. Mutations in the *FSH*β gene have been reported. These have an autosomal recessive pattern of inheritance and lead to low serum FSH and estradiol levels and high plasma LH levels. The clinical features include minimal development of secondary sexual characteristics and amenorrhea with no history of menses. Combined hormone deficiencies have also been described. Mutations in *Prop-1*, a pituitary transcription factor for Pit-1, lead to deficiencies in gonadotropins, TSH, PRL, and GH. These patients present with stunted growth, hypothyroidism, and delayed puberty in addition to amenorrhea.

B. Endocrine causes   *Hyperprolactinemia* is one of the most common causes of amenorrhea, accounting for 15% to 30% of cases. In the absence of pregnancy or postpartum lactation, persistently elevated PRL is almost always associated with a hypothalamic-pituitary disorder. Normal PRL secretion is regulated by several stimulatory and inhibitory factors (see Chapter 4). PRL secretion is primarily under tonic inhibition by dopamine, so that any interference with dopamine synthesis or transport from the hypothalamus may result in elevated PRL levels. In addition to menstrual disturbances, individuals with hyperprolactinemia may present with galactorrhea. In fact, hyperprolactinemia is a common cause of galactorrhea, and up to 80% of patients with amenorrhea and galactorrhea have elevated PRL levels. Other associated symptoms include headaches, visual field defects, infertility, and osteopenia.

The mechanism whereby hyperprolactinemia causes amenorrhea is not completely known. Studies have shown that PRL can affect the reproductive system in several ways. PRL receptors have been identified on GnRH neurons and may directly suppress GnRH secretion. Others have postulated that elevated PRL levels inhibit GnRH pulsatility indirectly by increasing other neuromodulators such as endogenous opioids. There is also evidence that GnRH receptors on the pituitary may be downregulated in the presence of hyperprolactinemia. Furthermore, PRL may affect the ovaries by altering ovarian progesterone secretion and estrogen synthesis. The best data now available suggest that hyperprolactinemia causes amenorrhea primarily by suppression of GnRH secretion.

Approximately half of patients with elevated PRL levels have radiologic evidence of a pituitary tumor. The most common type is a PRL-secreting tumor (prolactinoma), accounting for 40% to 50% of pituitary tumors. Prolactinomas are composed mainly of lactotrophs, and they secrete PRL. Occasionally these tumors secrete both GH and PRL.

The diagnosis of a pituitary adenoma is usually made by examination of the pituitary with magnetic resonance imaging (MRI). These tumors are categorized into two groups based on their dimensions—microadenomas are those less than 10 mm in diameter and macroadenomas are larger than 10 mm. These tumors are usually located in the lateral wings of the anterior pituitary. Microadenomas are typically wholly contained within the confines of the pituitary gland. Very rarely, a microadenoma infiltrates the surrounding tissue, including the dura, cavernous sinus, or adjacent skull base. A macroadenoma may expand farther and grow out of the sella to impinge on surrounding structures, such as the optic chiasm or may extend into the sphenoid sinus. As a result, macroadenomas are more frequently associated with severe headaches, visual field defects, and ophthalmoplegia. The incidence of a microadenoma progressing to a macroadenoma is relatively low, only 3% to 7%. During pregnancy the risk of a microprolactinoma enlarging is also low, but in the presence of a macroprolactinoma, the chance of tumor growth is up to 25%.

Some investigators have found a correlation between pituitary adenoma size and serum PRL levels. If the serum PRL is less than 100 ng/mL, a microprolactinoma is more likely, whereas if it is greater than 100 ng/mL, a macroprolactinoma is often present.

Although this correlation has been reported, the evidence supporting it is not strong. In fact, low PRL levels may also be associated with nonfunctioning macroadenomas. This is thought to reflect impingement of the adenoma on the pituitary stalk leading to reduced dopamine secretion and loss of the tonic suppression of prolactin secretion. These nonfunctioning tumors may synthesize glycoproteins such as FSH, LH, or their free alpha subunits. Rarely, functioning tumors may arise from other pituitary cells, resulting in excessive hormone secretion. If a macroadenoma is present, measurement of IGF-1, alpha subunit, TSH, and 24-hour urinary cortisol will begin to exclude other functioning adenomas.

Other tumors of nonpituitary origin may also result in delayed puberty and amenorrhea. The most common of these is a craniopharyngioma. Although craniopharyngiomas are commonly located in the suprasellar region, they can also involve the sella. These tumors have not been shown to produce hormones, but because they may compress the infundibulum, they can interfere with the tonic inhibition of PRL secretion and result in mildly elevated PRL levels.

Hyperprolactinemia in a patient with amenorrhea is defined as a PRL level greater than 20 ng/mL, although the upper limit of normal may vary between laboratories. Normal PRL release follows a sleep-circadian rhythm, but PRL may also be secreted in response to stress, physical exercise, breast stimulation, or a meal. Therefore, PRL should be measured in the midmorning hours and in the fasting state. Other causes of mildly elevated PRL include medications such as oral contraceptives, neuroleptics, tricyclic antidepressants, metoclopramide, methyldopa, and verapamil. Hyperprolactinemia has also been observed in several chronic diseases, including cirrhosis and renal disease. Furthermore, inflammatory diseases such as sarcoidosis and histiocytosis can infiltrate the hypothalamus or pituitary and result in hyperprolactinemia. Elevated PRL levels may be a physiologic response. During pregnancy, PRL levels may be two to four times baseline. With postpartum breastfeeding, the PRL level should be below 100 ng/mL after 7 days and below 50 ng/mL after 3 months. If a woman is not breastfeeding, PRL levels should return to baseline by 7 days postpartum.

Persistently elevated PRL levels may also be present in primary hypothyroidism. Approximately 40% of patients with primary hypothyroidism present with a minimal increase in prolactin (25-30 ng/mL), and 10% present with even higher serum levels. Individuals with primary hypothyroidism have an increase in thyrotropin-releasing hormone (TRH) from the hypothalamus, which stimulates TSH and PRL release and leads to hyperprolactinemia. Patients with longstanding primary hypothyroidism may eventually manifest profound pituitary enlargement due to hypertrophy of thyrotrophs. This mass effect with elevated PRL levels mimics a prolactinoma. Therefore, all patients with hyperprolactinemia should have their thyroid function investigated to exclude hypothyroidism as the cause.

Prolactinomas are the most common cause of persistent hyperprolactinemia. All patients with elevated PRL levels should have the test repeated for verification. In addition to blood tests, a careful clinical and pharmacologic history and physical examination

should be performed to exclude other causes of hyperprolactinemia. If elevated PRL levels persist or if any measurement is found to be above 100 ng/mL, MRI of the hypothalamic-pituitary region should be performed. If a microadenoma is observed, the diagnosis of microprolactinoma can be made. If a macroadenoma is observed, other pituitary hormones should be measured to exclude other functioning adenomas or hypopituitarism. All patients diagnosed with macroadenoma should have a visual field examination.

The treatment of choice for prolactinoma is dopamine agonist therapy. These drugs (bromocriptine, cabergoline, pergolide, quinagolide) are very effective at lowering PRL levels, resolving symptoms, and causing tumor shrinkage. Treatment results in a rapid reduction in PRL levels in 60% to 100% of cases. Following reduction of PRL levels, 60% to 100% of women resume ovulatory menses within 6 weeks, and galactorrhea disappears within 1 to 3 months after starting treatment. Reduction in tumor size is usually evident after 2 to 3 months of drug therapy, but it may occur within days after initiation of treatment. The extrasellar portion of the tumor appears to be particularly sensitive to drug therapy, which explains the improvement in symptoms such as visual impairment or ophthalmoplegia with drug therapy. In patients diagnosed with a microadenoma that is manifested only as a menstrual disturbance, observation should only be considered if the patient does not desire conception. These individuals should be offered oral contraceptives to control the bleeding pattern and to protect bone from estrogen deficiency. However, the long-term sequelae of persistently elevated PRL levels are unknown. If dopamine agonist treatment is initiated for a microadenoma, therapy may be continued long term. If the tumor responds, the dose may be tapered and stopped after menopause. Many individuals with a macroadenoma take dopamine agonist therapy indefinitely. All patients diagnosed with a prolactinoma should have follow-up imaging, determination of serum PRL levels, and visual field examinations. Alternatively, patients with a micro- or macroprolactinoma may be offered discontinuation of therapy. Dopamine agonist therapy may be tapered and ultimately discontinued if follow-up PRL levels are normal, and follow-up MRI scans show no evidence of tumor (or tumor reduction of at least 50%, with the tumor at a distance of more than 5 mm from the optic chiasm and no evidence for tumor invasion of adjacent structures), and if follow-up after withdrawal is assured.

An alternative to medical management of pituitary tumors is transsphenoidal surgery, after which resolution of symptoms may be immediate. However, the success and recurrence rates vary and are dependent on the size of the tumor and the depth of invasion. The larger and more invasive the tumor is, the less chance there is for complete resection and the greater the chance of recurrence. In general, the success rate of surgery for a microadenoma in some centers may be 70% or higher; surgical cure rates for macroadenomas are generally less than 40%. Overall, the recurrence rate following surgery is approximately 50%. Surgery is a good alternative for resistant tumors or for patients intolerant of medical treatment. Because non-PRL-secreting pituitary tumors often do not respond well to medical therapy, operation is the treatment of choice for these tumors as well. The risks of surgery include infection,

diabetes insipidus, and panhypopituitarism. Complete pituitary testing should be performed prior to surgery.

Women with hyperprolactinemia who desire to conceive should be administered dopamine agonist therapy. Elevated PRL levels can lead to anovulation and hormonal disturbances that can result in difficulty in achieving pregnancy. Once pregnant, patients with microadenomas may have their medication discontinued. If a macroadenoma is present, thought should be given to continuing dopamine agonist therapy throughout pregnancy. Women with prolactinomas may breastfeed since there is no evidence that it stimulates tumor growth. If breastfeeding, dopamine agonist therapy should be discontinued since these agents will interfere with lactation.

C. Anatomic causes

1. **Pituitary destruction**—Amenorrhea may be the result of not only pituitary neoplasms but also pituitary destruction. Infiltrative disorders that involve the hypothalamus can affect the pituitary. Other disorders that cause pituitary insufficiency may be situational. A rare autoimmune disease, lymphocytic hypophysitis, can cause pituitary destruction during the puerperium and ultimately results in panhypopituitarism. Pituitary necrosis may also occur secondary to a hypotensive event. Typically, 80% to 90% of the pituitary must be damaged before pituitary failure ensues, and the robust blood supply to the pituitary makes this an uncommon event. If pituitary ischemia and necrosis are related to postpartum hemorrhage, it is known as Sheehan syndrome; otherwise, it is called Simmond disease. Because this type of injury typically affects the entire pituitary, most often more than one or all the pituitary hormones may be deficient. Observation has suggested that hormone loss follows a pattern, starting with the gonadotropins, followed by GH and PRL; however, this is highly variable.

OVARIAN AMENORRHEA

The physiologic period in a woman's life when there is permanent cessation of menstruation and regression of ovarian function is known as menopause (see later). The cause of ovarian failure is thought to be depletion of ovarian follicles. The median age at menopause is 51.1 years. Premature menopause is defined as ovarian failure prior to age 40, which is reported to occur in 1% of the population. The etiology of premature ovarian failure (POF) may have a genetic basis. Several mutations that affect gonadal function have been identified and include defects in hormone receptors and steroid synthesis. Other potential causes include autoimmune ovarian destruction, iatrogenic ovarian injury, and idiopathic ovarian failure. The most severe form of POF presents with absent secondary sexual characteristics and is most often due to gonadal agenesis or dysgenesis (see Chapter 14). Less severe forms may result only in diminished reproductive capacity.

The other ovarian cause of amenorrhea is repetitive ovulation failure or anovulation. Other than menopause, it is the most common cause of amenorrhea. Chronic anovulation may be secondary to disorders of the hypothalamic-pituitary axis and has been

previously discussed. Anovulation may also be due to systemic disorders. The causes of ovarian failure and anovulation due to peripheral disorders will be discussed later.

Ovarian failure is diagnosed based on the clinical picture of amenorrhea and the demonstration of elevated FSH (>40 IU/L). This may occur at any time from embryonic development onward. If it occurs prior to age 40, it is called POF. The presence or absence of secondary sexual characteristics defines whether ovarian activity was present in the past. The most common cause of hypergonadotropic amenorrhea, in the absence of secondary sexual characteristics, is abnormal gonadal development, which occurs in more than half of these individuals. When the gonad fails to develop, this is known as **gonadal agenesis**. The karyotype of these individuals is 46,XX, and the cause of failure is usually unknown. If streak gonads are present, this indicates at least partial gonadal development and is called **gonadal dysgenesis**. The karyotype of these individuals may be normal, but it is more likely that there are alterations in sex chromosomes (see Chapter 14).

Premature Ovarian Failure

A. Genetic origin of POF   Two intact X chromosomes are necessary for the maintenance of oocytes during embryogenesis, and the loss of or any alteration in the sex chromosome leads to accelerated follicular loss. This implies that two intact alleles are required for the normal function of some genes on the X chromosome. Turner syndrome is a classic example of complete absence of one X chromosome. It manifests as short stature, sexual infantilism, amenorrhea, and ovarian dysgenesis. This is a well-recognized condition that occurs in 1:2000 to 1:5000 females at birth. Turner syndrome is associated with a number of other phenotypic abnormalities, including a webbed neck, broad chest, low hairline, and cardiovascular and renal defects. It is interesting that fewer than half of patients with Turner syndrome have a single cell line with the karyotype 45,X. The majority of patients actually present with a mosaic karyotype such as 45,X/46,XX. These patients have varying degrees of the Turner syndrome phenotype and may display some secondary sexual development or may have a history of menstrual function. Some pregnancies have been reported.

The term **mixed gonadal dysgenesis** is used to describe chromosomal aneuploidies with a Y-chromosome component. The most common type is a mosaic pattern that has been associated with Turner syndrome (45,X/46,XY). These patients may have some functional testicular tissue and present with varying degrees of genital ambiguity. If enough testicular tissue is present to produce AMH, these patients may also present with abnormalities of the internal genitalia. Individuals who have both ovarian tissue and testicular tissue along with Wolffian and Müllerian structures internally are termed **true hermaphrodites**. Interestingly, most true hermaphrodites have 46,XX karyotypes (see Chapter 14) but others may have a 46,XY or 46,XX/46,XY chimerism.

Patients with gonadal dysgenesis may be phenotypically normal, and the abnormality may be manifest only as delayed pubertal development and amenorrhea. They probably have normal Müllerian structures and streak gonads. These individuals can display an array of karyotypes, including 46,XY (Swyer syndrome).

Patients with a male karyotype but a female phenotype presumably underwent testicular failure prior to internal or external genitalia differentiation. If a dysgenetic gonad contains a Y chromosome or a fragment of the Y chromosome, there is a 10% to 30% risk for future gonadal malignancy, and the risk is higher if a mutation in the *SRY* gene is present. These tumors may be hormonally active. The onset of spontaneous pubertal development in girls with mixed gonadal dysgenesis may be a clinical marker of tumor development. Gonadal extirpation is performed at the time of diagnosis.

POF is defined as ovarian failure before age 40 but after puberty. Because complete absence of an X chromosome results in a dysgenetic gonad, candidate genes for POF are probably those that escape X inactivation. In mammals, X inactivation occurs in all cells in order to provide dosage compensation for *X-linked* genes between males and females (Lyon hypothesis). Further observation has illustrated that terminal deletions in Xp lead to the classic stigmata of Turner syndrome, whereas other deletions in Xp or Xq present with varying degrees of early reproductive failure. Most of the genes involved in folliculogenesis appear to be located on the long arm of the X chromosome. Several regions on the X chromosome, including POF 1 and POF 2, have been evaluated with knockout models in animals and have shown varying effects on ovarian development (Figure 13–12).

Limited observations have found that deletions occurring closer to the centromere of the X chromosome manifest a more severe phenotype that includes disruption of pubertal development. In contrast, deletions that occur in the distal regions tend to present with early reproductive aging and infertility. An example of a distal mutation on the long arm of the X chromosome is that in the *FMR1* gene (*fragile X gene*). An association has been
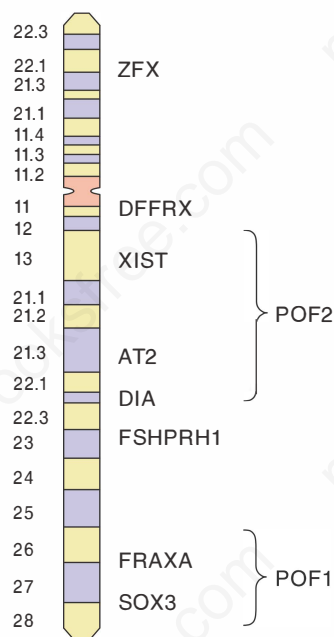
FIGURE 13–12 Candidate genes on the X chromosome for premature ovarian failure (POF). (Reproduced with permission from Davison RM, Davis CJ, Conway GS. The X chromosome and ovarian failure. *Clin Endocrinol (Oxf)*. 1999 Dec;51(6):673-679.)

described between the *FMR1* permutation state and POF. The prevalence of *FMR1* gene permutations accounts for approximately 2% to 3% of patients who present with sporadic POF and may be as high as 15% in familial cases. This permutation is also associated with the occurrence of a late-onset neurological disorder in male carriers designated as fragile-X tremor ataxia syndrome (FXTAS). Although a number of genes on the X chromosome have demonstrated involvement in ovarian physiology, the majority of patients with POF have no identifiable mutations on the X chromosome.

Autosomal recessive genes that have been shown to contribute to POF are very rare. FSH receptor mutations have been identified in humans with POF. These individuals present with a phenotype that ranges from absent secondary sexual development to normal development and early reproductive failure. The prevalence of FSH receptor mutations varies but is most common in the Finnish population (1% carriers). This mutation has not been observed in North America. An inactivated LH receptor has been identified in patients with normal puberty and amenorrhea but is quite rare. Mutations in genes involved in steroidogenesis have also been associated with POF. These enzymes include CYP17α and aromatase. Patients with CYP17α mutations may have a 46,XX or 46,XY karyotype. They have a similar phenotype except that those with 46,XY have absent Müllerian structures because AMH is produced from their testes. Individuals with aromatase deficiency present with sexual ambiguity and clitoromegaly. Several other autosomal mutations have been discovered that may have a role in ovarian physiology. However, at this time, most cases of POF with normal pubertal development have not been associated with any specific mutation (see Chapter 14).

B. Autoimmune origin of POF

Autoimmune ovarian destruction is another potential cause of POF. This diagnosis is difficult to make unless it presents with one of the autoimmune polyglandular syndromes (APSs) (see Chapter 2). The circumstantial evidence supporting the diagnosis is found in the high incidence of concomitant autoimmune disease—20% or more in patients with POF. The strongest association is with autoimmune thyroid disease. In addition, 10% to 20% of individuals with autoimmune adrenal disease experience POF. Conversely, 2% to 10% of patients with idiopathic POF develop adrenal insufficiency.

Autoimmune-associated POF is often diagnosed based on the presence of another autoimmune disease or detection of autoantibodies. Thyroid antibodies are most frequently screened. However, these antibodies are present in 15% to 20% of reproductive-aged women. Others have suggested measuring ovarian-specific autoantibodies. There is, however, significant inconsistency in the testing methodology, and this is not recommended. Although the diagnosis is imprecise, all patients suspected of having POF should be screened regularly for thyroid dysfunction. In addition, these patients should be evaluated for the APS and particularly tested for adrenal insufficiency, especially if there is no identifiable cause of POF. Testing for Addison disease can be performed using provocative ACTH testing or alternatively

adrenal antibodies to identify patients at risk. There are commercially available tests for both antiadrenal antibodies and antibodies to 21-hydroxylase. The sensitivity of these ACTH testing for the diagnosis of adrenal insufficiency is 100%, the specificity 98%, and the PPV is 67%. If a patient has either of these antibodies they have a 25% to 50% risk of developing adrenal insufficiency within 5 to 10 years.

C. Iatrogenic causes of POF

Iatrogenic causes of POF include radiation therapy, chemotherapy, and ovarian insults resulting from torsion or surgery. The risk of POF following radiation and chemotherapy is proportionate to the patient's age. If the radiation dose is higher than 800 Gy, all women experience ovarian failure. Displacing the ovaries out of the radiation field (ovarian transposition) has shown to be very effective at preserving ovarian function in patients receiving pelvic radiation. Therefore, patients wishing to preserve their fertility should be offered this procedure. Chemotherapy alone, particularly with alkylating agents, may induce temporary or permanent ovarian failure. Studies have shown that the likelihood of experiencing amenorrhea during treatment is 50% to 100%, and on average 40% of patients experience irreversible menopause. In general, younger individuals with chemotherapy-induced ovarian injury are more likely to recover. In women older than age 40, the chance of recovery is less 10%. Modifications of chemotherapy drugs, duration of treatment, and dosing are modifiers of permanent ovarian injury.

D. Resistant ovary syndrome

A rare cause of hypergonadotropic amenorrhea associated with numerous unstimulated ovarian follicles is the resistant ovary syndrome. These patients classically have no history of ovulatory dysfunction and present with secondary sexual characteristics and symptoms suggestive of estrogen deficiency. This diagnosis was established in an era when ovarian biopsy was used to determine the cause of menstrual disturbances. However, the definition of resistant ovary syndrome is not universally accepted. In fact, in the original series, cases were included in which patients had demonstrated ovulatory function in the past but later developed a clinical picture suggestive of ovarian resistance. This pattern is more typical of ovarian aging and follicular depletion.

The cause is not known. Histologic features of ovarian biopsy demonstrate that there is no plasma cell or lymphocytic infiltration, indicating that it is not caused by autoimmune destruction. The presence of numerous follicles indicates that POF is not due to follicular depletion. Several studies have looked at gonadotropins, FSH receptors, and antibodies that serve as blockers to the gonadotropin receptors, and the literature to date is inconclusive about the cause.

Diagnosis can be established with certainty only with ovarian biopsy. However, current recommendations for management of amenorrhea do not include surgery to make a diagnosis. The diagnosis is, therefore, one of exclusion. In the absence of autoimmune disease and of any history of ovulation, karyotyping should be performed to exclude chromosomal abnormalities. In patients with normal karyotypes, the diagnosis of POF and resistant ovary

syndrome is difficult to make without biopsy. Improvements in ultrasound technology may make it possible to differentiate these entities by the measurement of ovarian volumes and antral follicle counts.

E. Summary   In more than half of patients with POF, no specific cause can be identified. The age defining POF is somewhat arbitrary. By definition, menopause is preceded by reproductive failure. It is thought that the time interval between menopause and the end of fertility may be approximately 10 years, and we know that approximately 10% of women reach menopause by 46 years of age and 1% by 40 years. Therefore, women who experience menopause at 45 years of age probably encounter a decline in reproductive potential or even reproductive failure at 35 years of age. This has obvious implications for women who are delaying childbearing. Several studies have described a significant association between the menopausal ages of mothers and daughters, twins, and sisters. A number of studies have identified new genes that are involved in ovarian physiology. It is hoped that these investigations will help in the treatment of subfertility and result in a reduction in infertility. At a minimum, they may allow better prospective individual prediction of reproductive risk.

Women diagnosed with POF that is not readily explained must have an evaluation that consists of a karyotype, testing for fragile-X (FMR) permutation, and evaluation for APS. If a patient has an FMR permutation, one should offer screening to her father. If there is no identifiable cause of POF, screening for adrenal insufficiency is obligatory. All patients with POF should be considered for combination HT. The increased risks associated with breast cancer are likely no greater than the age-adjusted risk of women with intact ovaries. Patients with absent secondary sexual characteristics should initially be given estrogen-only therapy in low doses and titrated every 3 to 6 months (see Chapters 14 and 15 and see later for management of menopause).

ANOVULATION

Chronic anovulation may be defined as repetitive ovulation failure, which differs from ovarian failure in that viable oocytes remain in the ovary. Anovulation is the most common cause of amenorrhea during the reproductive years. There are several causes; those associated with hypothalamic and pituitary disorders have previously been mentioned and will not be considered in this section. Other conditions that cause anovulation include the peripheral endocrinopathies. These disorders result in a hormonal imbalance—mainly elevated androgens or estrogens—and lead to inappropriate feedback mechanisms and ovulatory failure. The peripheral endocrine disorders will be discussed later in greater detail.

Hyperandrogenism and Anovulation

A. Polycystic ovarian syndrome   Hyperandrogenic anovulation accounts for more than 30% of cases of amenorrhea and up to 75% of all cases of anovulation. Most often it is due to polycystic ovarian syndrome (PCOS). The reported prevalence of PCOS depends on the criteria used to define it. Although there remains considerable controversy over the definition, most investigators, until recently, have focused on the 1990 National Institutes of Health-National Institute for Child and Human Development (NIH-NICHD) diagnostic criteria (Table 13–4). This definition puts forth ovarian hyperandrogenism and anovulation as the cardinal features of PCOS. The criteria include ovulatory dysfunction with evidence of hyperandrogenism, either clinically or by laboratory testing, in the absence of identifiable causes of hyperandrogenism. Using these criteria, the prevalence of unexplained hyperandrogenic chronic anovulation approximates 4% to 6%, and it is considered the most common endocrine disorder in women of reproductive age. However, in 2003 the diagnostic criteria were revised by the Rotterdam consensus conference (see Table 13–4). The Rotterdam criteria consider a broader spectrum of ovarian dysfunction than previous definitions. Some experts criticized the definition, because it encompasses many more women, notably those with hyperandrogenism and ovulatory cycles, in addition to women with anovulation without androgen excess. This is contrast to the 1990 NIH-NICHD definition, which required *both* androgen excess and irregular cycles. Most recently, the Androgen Excess Society task force proposed the criteria should be defined by the presence of hyperandrogenism and ovarian dysfunction (see Table 13–4). This definition expands upon the 1990 NIH-NICHD criteria and requires that all patients with PCOS are hyperandrogenic at some level. Although the revised criteria can include women with ovulatory cycles, the remaining discussion on PCOS focuses on women who have anovulatory cycles (Table 13–5).

Approximately 50% of women diagnosed with PCOS are obese, and most have polycystic ovaries present on sonography (see later). Underlying these features are numerous biochemical abnormalities that have been associated with this syndrome, including elevated circulating total testosterone, free testosterone, DHEAS, and insulin as well as decreased SHBG and an elevated LH-FSH ratio. However, these abnormalities are not present in all PCOS patients. In fact, 40% of women who present with only hirsutism (see later) have elevated total testosterone levels, and

TABLE 13–4 Revised diagnostic criteria for polycystic ovary syndrome.

1990 criteria (both 1 and 2)
1. Chronic anovulation
2. Clinical and/or biochemical signs of hyperandrogenism and exclusion of other etiologies
Revised 2003 criteria (2 of 3)
1. Oligo or anovulation
2. Clinical and/or biochemical signs of hyperandrogenism
3. Polycystic ovaries and exclusion of other etiologies (congenital adrenal hyperplasia, androgen-secreting tumors, Cushing syndrome)

Reproduced with permission from Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril.* 2004 Jan;81(1):19-25.

TABLE 13–5 All possible phenotypes based on the presence or absence of oligo-anovulation, hyperandrogenemia, hirsutism, and polycystic ovary syndrome.

Potential Phenotypes																
Features	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Hyperandrogenemia	+	+	+	+	−	−	+	−	+	−	+	−	−	−	+	−
Hirsutism	+	+	−	−	+	+	+	+	−	−	+	−	−	+	−	−
Oligo-anovulation	+	+	+	+	+	+	−	−	−	+	−	−	+	−	−	−
Polycystic ovaries	+	−	+	−	+	−	+	+	+	+	−	+	−	−	−	−
NIH 1990 criteria	✓	✓	✓	✓	✓	✓										
Rotterdam 2003 criteria	✓	✓	✓	✓	✓	✓	✓	✓	✓							
AE-PCOS 2006 criteria	✓	✓	✓	✓	✓	✓	✓	✓								

Abbreviations: AE, androgen excess; PCOS, polycystic ovarian syndrome.

Reproduced with permission from Azziz R, Carmina E, Dewailly D, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil Steril.* 2009 Feb;91(2):456-488.

30% to 70% have elevated DHEAS levels. Similarly, the evaluation of increased LH pulsatility, in association with low-normal FSH levels (LH-FSH ratio), is not a reliable diagnostic test. Although elevated LH-FSH ratios are common findings in thin women, in obese patients with PCOS, the ratio is within the normal range about half of the time. The short half-life of LH (~20 minutes) is likely another major contributor to the inaccuracy of LH testing. Hyperinsulinemia has recently been hypothesized to play a major role in the pathogenesis of PCOS (see later). The prevalence of insulin resistance may approximate 50% to 60%, compared with 10% to 25% observed in the general population. However, insulin resistance is difficult to measure in the clinical setting. Part of the difficulty is that there is no universally agreed on definition of insulin resistance, and the laboratory tests are not standardized. Furthermore, baseline insulin levels vary depending on population and body weight. For example, up to 60% of ovulatory obese patients have demonstrated some form of insulin resistance. Nonetheless, there is good evidence that a subset of normal-weight women and obese women with PCOS have a greater degree of insulin resistance and compensatory hyperinsulinemia compared with weight-matched controls.

There is increasing evidence for a strong genetic component in the etiology of PCOS. Several candidate genes have been investigated, including genes involved in steroidogenesis and carbohydrate metabolism, but none has been conclusively linked with the disease. PCOS is heterogeneous clinically, raising the possibility of different genetic causes and a variable environmental contribution to the syndrome.

1. **Diagnosis of PCOS**—The diagnosis of PCOS is typically based on clinical features (irregular menstrual cycles, acne, hirsutism), although additional information may be obtained with biochemical testing and sonographic examination. Known causes of hyperandrogenism and anovulation should be excluded in all patients (ie, androgen-secreting tumors, thyroid and adrenal gland dysfunction, hyperprolactinemia).

In most situations, the manifestations of PCOS emerge in the peripubertal years, occasionally with premature pubarche, and likely irregular cycles that persist through much of reproductive life. The diagnosis may be challenging, however, because adolescent girls commonly display irregular cycles during the 5 years following menarche. Ultrasonography may help to solidify the diagnosis of PCOS in children given that the likelihood of restoration of normal ovarian morphology is small as long as irregular menses persist.

Polycystic ovaries tend to be enlarged and have been defined in multiple ways: (1) the presence of 10 or more cystic follicles that are between 2 and 8 mm in diameter and arranged along the subcapsular edge of the ovary in a string of pearls fashion; and (2) 12 or more follicles 2 to 9 mm in either ovary and/or an ovarian volume of 10 cm^3 or greater. The increased ovarian volume and displacement of follicles toward the periphery may be explained by the hyperplastic stroma. The stroma contains hilar cells and secondary interstitial cells (theca cells) that are capable of androgen synthesis.

Polycystic-appearing ovaries (PCAO) are very common in women with clinical features suggestive of hyperandrogenism, independent of menstrual disturbances. This finding suggests that ovarian hyperandrogenism can occur in ovulatory states and should be considered in the spectrum of PCOS, if polycystic ovaries are present. However, the finding of polycystic ovaries alone does not establish the diagnosis of PCOS. In fact, based on the Rotterdam definition, more than 30% of normal women have this ovarian morphologic feature. The prevalence is significantly impacted by age. Approximately 60% of women with regular cycles between the ages of 25 to 30 have morphologic characteristics consistent with polycystic ovaries compared to 7% of women between the ages of 41 and 45. Furthermore, the morphogenesis of polycystic ovaries is not unique to PCOS as it has been observed in other scenarios (ie, late-onset congenital adrenal hyperplasia, HIV, epilepsy). Given the prevalence of PCOA, a new definition has been proposed: PCAO or polycystic ovarian morphology (PCOM) is a follicle count of more than or equal to 25 within one ovary, but only when using newer technology that affords maximal resolution of ovarian follicles (ie, transducer frequency ≥8 MHz). If such technology is not available, ovarian volume cutoff of more than 10 cm^3 is suggested. This definition is supported by a threshold, when discriminating between controls and PCOS, of 26 follicles having the best compromise between sensitivity (85%) and specificity (94%) compared to 69% sensitivity and 90% specificity with a follicle count cutoff of 12 or a 81% sensitivity and 84% specificity with ovarian volume (10 cm^3).

2. **Hyperandrogenism**—The origin of sex steroids is from two sources: the gonads and the adrenal gland. The adrenal gland is responsible for most of the precursor hormones in the general circulation, which serve as reservoirs for the more potent androgens and estrogens. The ovary produces bioactive androgens and estrogens as well as precursor hormones. Figure 13–13 illustrates the relative contribution of each organ to the pool of circulating sex steroids. DHEAS is the most abundant steroid in the circulation, and serves as a precursor to the more potent androgens and estrogens. Over 98% of
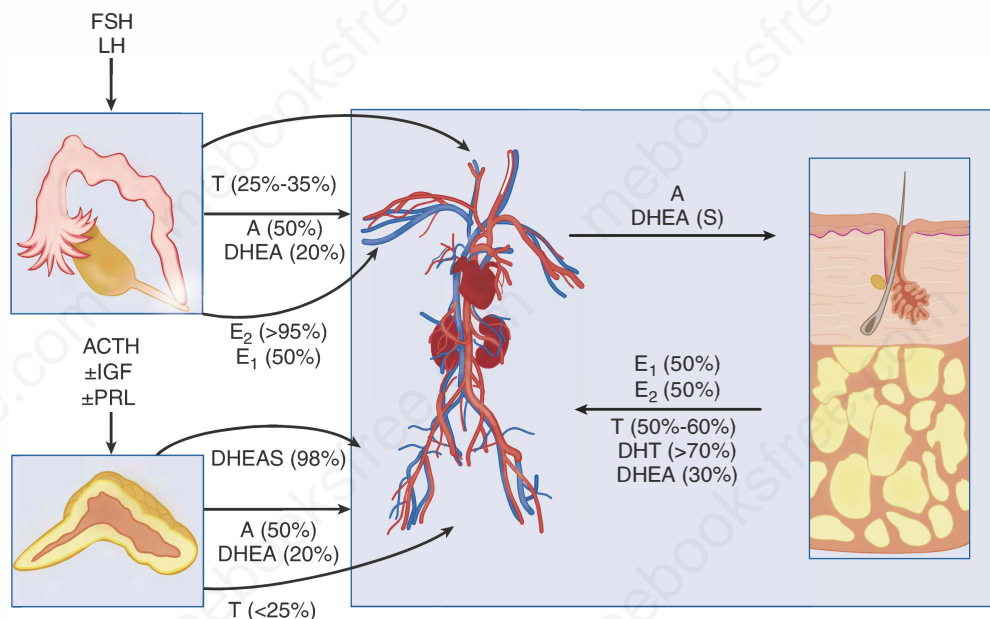
FIGURE 13–13 The origin of circulating sex steroids: the relative contribution of each organ to the pool of circulating sex steroids.

DHEAS is secreted from the adrenal gland. Approximately half of the circulating levels of androstenedione are equally derived from the adrenal gland and gonads and the remaining half from conversion of DHEAS (DHEA) in the periphery. DHEAS also serves as a substrate for the ovary to produce bioactive sex steroids. It is evident that the majority of bioactive androgen is derived from precursor steroids that undergo peripheral conversion. It is suggested over 60% of the testosterone is derived from circulating androstenedione, and the remaining is by direct secretion. The contribution of testosterone from the ovary and the adrenal gland is not completely known. The fact that the various methods to obtain secretion rates are difficult and that assays used to measure testosterone can be unreliable with resultant variability contributes to the uncertain interpretation.

A common clinical sign of hyperandrogenism is hirsutism. Hirsutism affects approximately 5% to 15% of the population. The prevalence depends on the population and the method used to establish the diagnosis. The most common scoring method used to diagnose hirsutism is based on the Ferriman and Gallwey system. It is based on evaluating 9 to 11 body areas and assigning a score of 0 to 4 based on the density of hair. This method is primarily used for research purposes. Clinically, it is often diagnosed by history and the presence of excess hair growth in centrally located regions, not commonly found in women. For example, those areas most affected are the face (sideburns, mustache, and beard), chest, linea alba, or the inner aspect of the thighs. The characteristics and distribution of body hair may be influenced by ethnic or racial factors. Alternatively, hyperandrogenism may present as acne or alopecia.

The diagnostic evaluation of hirsutism should include a thorough history and physical examination (Figure 13–14). The etiology can be divided between non-androgenic and androgenic causes. Organic causes of hyperandrogenism must be excluded. Non-androgenic causes include chronic skin irritation, anabolic medications, and rarely acromegaly. The most common androgenic cause is PCOS which affects at least 70% of hirsute women. Less common is late-onset congenital adrenal hyperplasia, which affects approximately 5% of patients. PCOS is a diagnosis of exclusion—measurements of serum androgen levels should play only a limited role in the evaluation. Most patients who have hyperandrogenemia present with obvious clinical manifestations, and the presence of normal androgen levels in a patient with hirsutism or acne does not exclude the diagnosis of PCOS. However, there are subsets of amenorrheic patients (ie, Asians) who are hyperandrogenemic without clinical manifestations, most likely as a result of their relative insensitivity (due to low 5α-reductase activity) to circulating androgens. It is in these patients that assessing androgen levels may be of value in determining the cause of amenorrhea. More commonly, androgens such as DHEAS and testosterone are measured to exclude other causes of hyperandrogenic anovulation such as nonclassic adrenal hyperplasia and androgen-secreting tumors (see later).

Another cause of hirsutism, made only after excluding other diagnoses, is characterized by regular menstrual cycles, normal androgen levels, and the absence of PCAO; it is termed idiopathic hirsutism. This common disorder occurs more frequently in certain ethnic populations, particularly those of Mediterranean descent. It is thought that the etiology of hirsutism in these patients is related to higher 5α-reductase activity (increased sensitivity) within the pilosebaceous unit.

3. **Mechanism of anovulation**—The mechanism of anovulation in PCOS remains unclear. It is evident that the population of antral follicles is increased and that follicular development is arrested. It is also known that the development of preantral follicles is not primarily under hormonal control. Evidence supports the components of the intraovarian network as regulators of antral follicle development. It is known that many of the accumulated follicles in PCOS remain steroidogenically
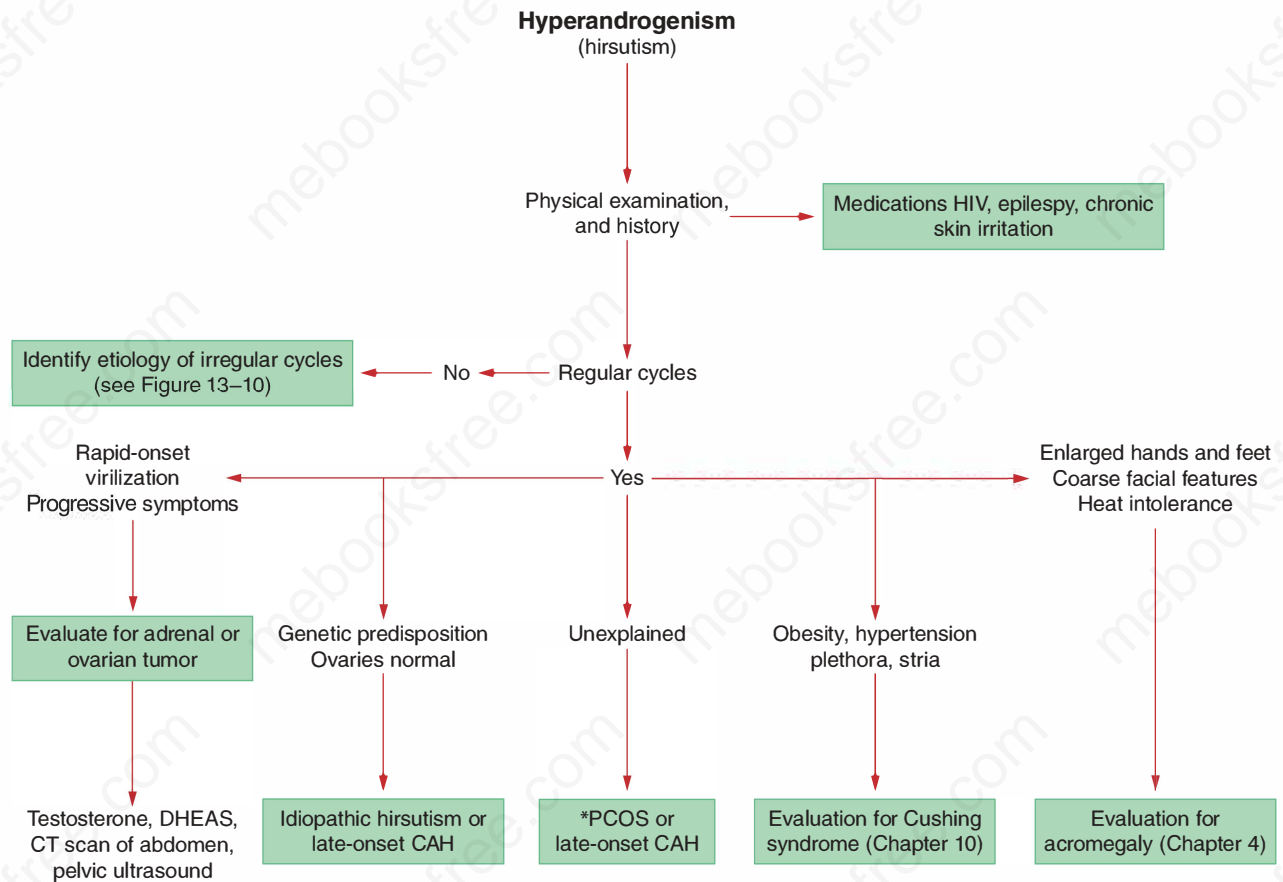
FIGURE 13–14 An algorithm for women with hirsutism. All women should be screened for known causes of hyperandrogenism. Hyperandrogenism suggests either PCOS or late-onset congenital adrenal hyperplasia (CAH). If hirsutism is rapid in onset or rapidly progressive in nature or the patient is frankly virilized, an evaluation for an androgen-secreting tumor is required. Management of hirsutism in late-onset CAH or PCOS is similar. *Late-onset CAH can present with regular cycles. Depending on the specific criteria (see Table 13–4), PCOS can present with regular cycles. Serum hormones are not required for the diagnosis of idiopathic hirsutism or PCOS.

competent and are capable of producing estrogen and progesterone. In fact, it is interesting that women with PCOS produce both androgens and estrogen (estrone) in excess.

One of the most frequently described characteristics of PCOS is the functional derangement of LH secretion. Numerous studies have shown that frequency, amplitude, and mean levels of LH are increased. The aberration in LH secretion may be a result of heightened pituitary responsiveness and increased hypothalamic GnRH activity. Under normal conditions, follicles respond to LH after they reach approximately 10 mm in diameter. However, polycystic ovarian follicles acquire responsiveness to LH at a much smaller diameter, which may lead to inappropriate terminal differentiation of granulosa cells and result in disorganized follicular development. It has been suggested that the theca cells increase their expression of steroidogenic enzymes following stimulation by LH, whereas the granulosa display resistance to FSH (see later). The elevated LH levels and relative hyperinsulinemia that exist in some PCOS patients may synergistically potentiate disordered folliculogenesis. Although hyperandrogenism is part of the diagnostic criteria for PCOS, its direct impact on folliculogenesis is not clear. It is conceivable that androgens contribute to the effects of LH and insulin on follicular maturation. It is also possible that the excess estrogens may result in a negative-feedback loop to inhibit FSH release and prevent further follicular development.

Most experts are of the opinion that excess androgen production is a fundamental abnormality in women with PCOS. Androgens within the ovary are produced mainly by the thecal interstitial cells that surround the follicle and to a lesser extent by the secondary interstitial cells located in the stroma (see earlier). The CYP17α complex is thought to be the key enzyme in biosynthesis of ovarian androgens. Under normal conditions, a large proportion of the androgens produced by the theca cells diffuses into the granulosa cell layer of the follicle where they are rapidly converted to estrogen as shown in Figures 13–9 and 13–13 (two-cell theory). The intrinsic control of androgen production in the ovary is modulated by intraovarian factors and hormones (see section on ovarian steroidogenesis at the beginning of this chapter). It is the dysregulation of hormone production that is most likely responsible for PCOS.

Several studies have shown that women with PCOS have an exaggerated ovarian androgen response to various stimuli. To illustrate, hyperstimulation of 17-hydroxyprogesterone levels was noted when women diagnosed with PCOS were given a

GnRH agonist or hCG, suggesting increased CYP17α activity. This study is supported by *in vitro* studies in which measurement of steroids in cultured human theca cells from polycystic ovaries revealed concentrations of androstenedione, 17α-hydroxyprogesterone, and progesterone that were respectively 20-fold, 10-fold, and 5-fold higher than levels in control cells. Additional studies have found increased expression of the genes encoding CYP17α hydroxylase, P450scc, the LH receptor, and StAR. These findings reflect a global enhancement of steroidogenesis. This situation is compounded by the hypertrophy of theca cells that is present in women with PCOS.

Several studies have evaluated intraovarian modulators as participants in the pathogenesis of PCOS (Figure 13–15). IGF-binding proteins (IGFBPs), especially IGFBP-2 and IGFBP-4, are found to be increased in the follicular fluid of polycystic ovaries. They may act locally to decrease free IGF-2 and, thus, decrease the effects of FSH on the oocyte and granulosa cells. Alternatively, the granulosa cells may downregulate their insulin receptors as a result of hyperinsulinemia. This would deprive the granulosa cells of their co-gonadotropin, IGF-2, and account for the relative FSH insensitivity. Inhibin is also a likely candidate, because a large proportion of women with PCOS have relative FSH suppression (see Figure 13–15). However, studies have not shown consistent results, suggesting that if inhibin is involved, the effect is minimal. Follistatin, the activin-binding protein, was in the past thought to play an important role in the development of PCOS. It was initially implicated because activin functions to inhibit androgen production and enhance FSH expression. However, current studies do not demonstrate a significant association between abnormalities in follistatin and PCOS.

The adrenal gland may be significantly involved in the pathogenesis of some cases of PCOS. The connection seems plausible because adrenal androgens can be converted to more potent androgens in the ovary. Furthermore, a significant portion of women with congenital adrenal hyperplasia have polycystic ovaries (see later). Several studies have shown that DHEAS is elevated in 25% to 60% of patients with PCOS. However, ACTH levels are normal in PCOS women. Interestingly, it has been reported that there is an increased response of androstenedione and 17α-hydroxyprogesterone to exogenous ACTH. These findings suggest an underlying abnormality in the CYP17α expressed in the adrenal gland as well as in the ovary. However, minimal data support CYP17α dysfunction in the adrenal gland. It has also been shown that ovarian steroids can stimulate adrenal androgen production; however, additional findings suggest that the ovary is not the primary cause of adrenal hyperresponsiveness. The critical role played by adrenal androgens during the pubertal transition has not been fully investigated as a potential contributor to the development of PCOS.

4. **Hyperinsulinemia and PCOS**—A relationship between insulin and hyperandrogenism has been postulated based on several observations. This association may be demonstrated in the peripubertal years with premature pubarche, which is more commonly associated with insulin resistance than with congenital adrenal hyperplasia or androgen excess. Various case
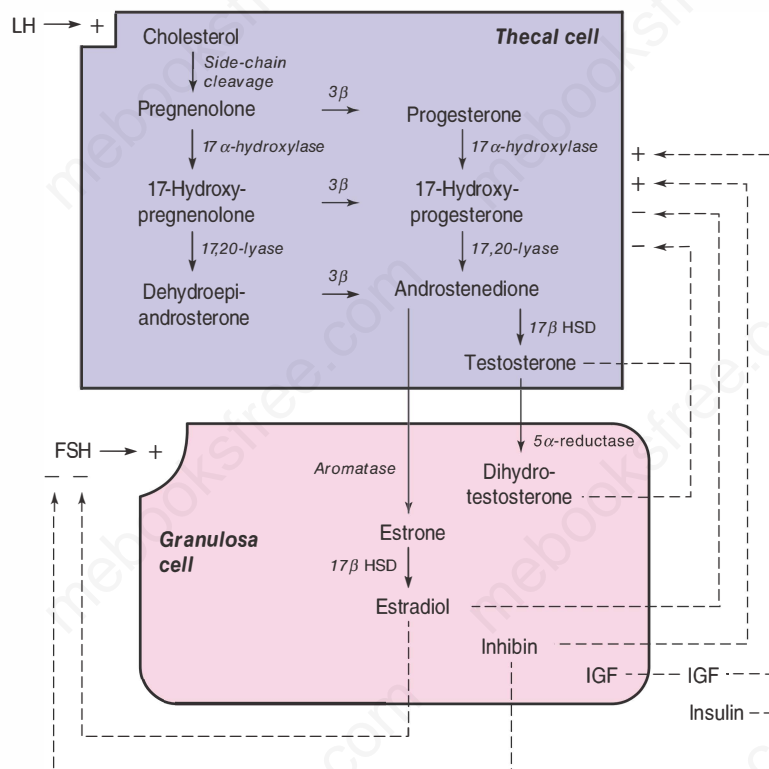
FIGURE 13–15 Regulation of sex steroid production in the theca and granulosa cells of the ovary. Several autocrine and paracrine factors such as insulin and insulin-like growth factors (IGFs) work synergistically with luteinizing hormone (LH) to stimulate androgen production through increased 17α-hydroxylase activity. (Reproduced with permission from Ehrmann DA, Rosenfield RL, Barnes RB, et al. Detection of functional ovarian hyperandrogenism in women with androgen excess. *N Engl J Med*. 1992 Jul 16;327(3):157-162.)

reports have shown that **acanthosis nigricans**—hyperpigmentation of skin in the intertriginous areas—is associated with severe insulin resistance. A number of these patients also presented with hyperandrogenism and anovulation. The relationship was substantiated when it was observed that the degree of hyperinsulinemia was correlated with the degree of hyperandrogenism. Further studies revealed that hyperinsulinemia is frequently identified in women with PCOS.

It has been shown that the cause of hyperinsulinemia is insulin resistance and that the defect lies in the postreceptor signaling pathway. The frequency and degree of hyperinsulinemia in women with PCOS is amplified in the presence of obesity. Although many women with PCOS exhibit insulin resistance, some do not. However, insulin resistance is also observed in some thin PCOS patients.

Insulin may cause hyperandrogenism in several different ways, although the exact mechanism has not been well defined. It has been suggested that insulin has a stimulatory effect on CYP17α. There is evidence from *in vitro* models that insulin may act directly on the ovary (see Figure 13–15). It has been shown that the ovary possesses insulin receptors and IGF receptors. In addition, several studies have reported that insulin stimulates ovarian estrogen, androgen, and progesterone secretion and that its effects are greatly enhanced by the addition of gonadotropins. Administration of an insulin-sensitizing agent (eg, metformin or a thiazolidinedione) to obese women with PCOS leads to a substantial reduction in 17α-hydroxyprogesterone levels, reflecting decreased CYP17α activity. However, clinical studies, in which insulin infusions were administered to normal women, failed to demonstrate increased testosterone production, and there were no changes in androgen levels when normal women were given insulin-sensitizing agents. These observations suggest that insulin has more of a modifying rather than a predisposing effect on androgen production.

The relationship between insulin and adrenal androgen production is less clear. Some studies have shown that insulin increases secretion of 17α-hydroxyprogesterone and DHEAS in response to ACTH. Other studies have shown that DHEAS decreases after acute insulin infusions are administered to normal men and women. Furthermore, when insulin-sensitizing agents were administered to women with PCOS, a decrease in DHEAS was observed. Although there is less evidence to support the association of insulin and adrenal androgen production, if there is an insulin effect, it is as a modulator of adrenal secretory activity.

Insulin may indirectly affect androgen levels. Several studies have reported that insulin directly inhibits SHBG production. There is an inverse correlation between insulin levels and SHBG, so that decreasing insulin levels would decrease the circulating bioavailable androgen level (via increases in SHBG). It has also been shown that insulin decreases IGFBP-1. This would increase free IGF-1, which could modulate ovarian androgen production in a fashion similar to insulin (see Figure 13–15). Although these indirect mechanisms may play a role, the literature suggests that insulin acts directly to augment androgen production (see Figure 13–15). However, it appears that a dysregulation in steroidogenesis must also exist in order for insulin to cause hyperandrogenism.

Increasing attention is being directed to the possibility that PCOS begins before adolescence. In fact, the initial insult may begin in utero, where there is ample exposure to androgens derived from the fetal adrenal gland and ovary. This hormonal environment may reprogram the ovary and alter steroidogenesis in a manner that predisposes to PCOS. The phenotypic expression of PCOS would then be determined by environmental factors such as diet and exercise.

The various biochemical abnormalities associated with PCOS have led to studies investigating metabolic sequelae of this syndrome. Long-term health problems, such as the development of cardiovascular disease and diabetes, have been linked to PCOS. In fact, several studies have suggested that the metabolic syndrome is significantly more prevalent in women with PCOS. However, studies looking at whether women with PCOS actually experience increased cardiovascular events are limited. Several observational studies have demonstrated that women with PCOS have alterations in their lipid profiles, including increased triglycerides and LDL and decreased high-density lipoprotein (HDL) compared with weight-matched controls. Furthermore, the degree of dyslipidemia has been correlated with the magnitude of insulin resistance. The fact that insulin resistance occurs with greater frequency in women with PCOS suggests that they are at higher risk for the development of diabetes mellitus. It is known that up to 30% to 40% of women with PCOS have impaired glucose tolerance, although it is most often seen in patients who are obese. Limited retrospective studies have suggested that women with PCOS have an increased frequency of developing type 2 diabetes. In summary, as demonstrated by surrogate markers, the available data show that these patients have increased risk factors for cardiovascular events and diabetes. Long-term prospective studies will be required to determine if these patients are actually at risk for increased mortality and morbidity.

The fact that hyperinsulinemia underlies many of the potential adverse sequelae raises a question about whether insulin resistance should be assessed in all patients diagnosed with PCOS. There are several methods of measuring insulin resistance (Table 13–6). However, the potential impact of hyperinsulinemia is unknown. Furthermore, there is no universal laboratory criterion or standardization for establishing the diagnosis of insulin resistance, which raises doubts about whether any potential adverse sequelae can be prevented if insulin resistance is identified. Further data are necessary, prior to widespread use of medications such as metformin or a thiazolidinedione to reduce insulin resistance in patients with PCOS to determine (1) if these patients are indeed at risk for

TABLE 13–6 Measurements of insulin resistance.

Frequently sampled intravenous glucose tolerance test (FSIGT)[a]
Euglycemic insulin clamp[a]
Insulin tolerance test (ITT)
  Fasting insulin levels
Oral glucose tolerance test (OGTT) + insulin levels and calculating area under the curve
Glucose-insulin ratio
  Derivatives of glucose-insulin ratio[b]
Insulin resistance index = glucose × insulin/25
Homeostasis model assessment of insulin resistance HOMA-IR = 22.5 × 18/glucose × or log (HOMA-IR)
Quantitative insulin sensitivity check index (QUICKD = 1/log (glucose) + log (insulin)

[a]American Diabetes Association (ADA) gold standard; not practical for clinical use.
[b]Calculations based on fasting glucose and insulin levels.

cardiovascular events, (2) if hyperinsulinemia is an independent risk factor, and (3) whether long-term treatment with any insulin-sensitizing agent will decrease the potential sequelae.

5. **Additional risks associated with PCOS**—Another concern is that women who are anovulatory do not produce a significant amount of progesterone. This leads to a situation in which the uterine lining is stimulated by unopposed estrogens, which is a significant risk factor for development of endometrial cancer. In fact, an association has been found between endometrial cancer and PCOS. There is also evidence to suggest an association of PCOS with both breast cancer and ovarian cancer, but PCOS has not been conclusively shown to be an independent risk factor for either disease.

6. **Management of PCOS**—The management of PCOS should be dictated by the patients' risk factors for cardiovascular disease, diabetes, malignant sequelae, and symptomatology (see also "Infertility"). All patients should have a lipid profile measured. In addition, oral glucose tolerance tests should be performed—at least in obese women. In patients who have a long history of irregular cycles (>1 year), an endometrial biopsy should be considered.

Cardiovascular risk factors, weight loss, and progression to diabetes can be improved with diet and exercise. In the subset of patients with glucose intolerance, administration of insulin-sensitizing agents should be considered. Metformin, a biguanide, is the most widely studied. Several studies have shown modest transitory weight loss, and possibly a decreased incidence of diabetes in patients with PCOS due to its use. Metformin is not effective for treatment of hyperandrogenic symptomatology. The use of metformin is not recommended unless there is evidence of metabolic derangements that are not fixed with diet and exercise alone.

Irregular bleeding can be improved by the administration of oral contraceptives which induce scheduled withdrawal bleeding. However, another benefit of oral contraceptives has been the significant reduction in endometrial cancer in the general population. It is rational to expect that a similar benefit would accrue to women with PCOS, because the ovarian stimulation is minimized and the progestin would counteract the estrogenic environment.

Acne and hirsutism in women diagnosed with PCOS or of idiopathic origin can be treated with oral contraceptives. The mechanism is not completely known, but oral contraceptives decrease the amount of bioavailable androgens by increased SHBG production and by ovarian suppression. It has also been shown that progestins can inhibit 5α-reductase activity, which further decreases the production of dihydrotestosterone, the major androgen that stimulates hair growth. The maximal effect is evident after 6 months of treatment (the hair cycle length is estimated to be 4 months).

If hirsutism is severe or if oral contraceptives alone are not effective, the addition of spironolactone may be beneficial. Spironolactone is an antimineralocorticoid agent that inhibits androgen biosynthesis in the adrenal gland and ovary, inhibits 5α-reductase, and is a competitive inhibitor of the androgen receptor. Side effects are minimal and include diuresis in the first few days; hyperkalemia, dyspepsia; nausea; skin hypersensitivity;

breast tenderness; and abnormal bleeding, which can be alleviated by increasing the dose to desired amount over 3 weeks. Spironolactone is used concomitantly with oral contraceptives. Because oral contraceptives and spironolactone act by different mechanisms, the combined effect is synergistic. It is important to note spironolactone can cause hyperkalemia and should be avoided in women with renal compromise. Additionally spironolactone is teratogenic. Therefore, concomitant contraception treatment should be administered in women of reproductive age. The only topical agent approved for hirustism by the FDA is Vaniqa (Eflornithine) cream, which is an inhibitor of L-ornithine decarboxylase. It has been shown to be effective for control of facial hair. Similar to spironolactone, Eflornithine is also teratogenic, and therefore concomitant contraception treatment should be administered in women of reproductive age. Long-acting GnRH agonists and 5α-reductase inhibitors (finasteride) have been used for refractory cases with some success. Shaving, bleaching, chemical depilation, plucking, or waxing is temporary measure to control unwanted hair growth. However, several of these methods can cause skin irritation and result in progressive hair growth. Permanent techniques such as electrolysis or laser depilation have shown promising results.

B. Congenital adrenal hyperplasia   Congenital adrenal hyperplasia is another disorder that may cause hyperandrogenism. It presents in a wide range of clinical forms, ranging from severe—which may be classified as *classic, salt-wasting,* or *simple virilizing*—to milder forms known as *acquired, adult-onset, nonclassic,* or *late-onset* congenital hyperplasia. The clinical manifestations reflect the severity of the enzymatic defect. Severe or classic forms are discussed in Chapter 14 and will not be considered here. This section will discuss the nonclassic forms, which affect 1% to 10% of the population, depending on the ethnicity of the patient. The clinical features are similar to those of patients diagnosed with PCOS and include menstrual irregularities, hyperandrogenism, and infertility, and up to 50% have polycystic ovaries.

The adrenal gland consists of a cortex and a medulla. The cortex is divided into three functional zones based on location and the principal hormone secreted (see Chapter 9). The zona glomerulosa is the outermost zone and lies adjacent to the adrenal capsule. It is primarily responsible for aldosterone production. The zona fasciculata lies immediately below the glomerulosa. It principally secretes glucocorticoids, although it is capable of producing androgens. The zona reticularis is located beneath the zona fasciculata and overlies the adrenal medulla. It is this zone that principally secretes androgens. Both the zona fasciculata and the zona reticularis are regulated by ACTH. It is the secretory activity of these two zones that results in nonclassic adrenal hyperplasia.

Congenital adrenal hyperplasia is an autosomal recessive disorder that is caused by mutations of genes involved with adrenal steroidogenesis. The mutations mostly occur in the 21-hydroxylase gene (*P450c21B*) and rarely in the 3β-hydroxysteroid dehydrogenase gene or 11β-hydroxylase genes (*P450c11B* and *P450c11AS*). In classic forms, the enzymatic defects are severe, affect both alleles, and result in cortisol deficiency diagnosed at birth. Owing to the cortisol deficiency, there is ACTH excess and

hyperstimulation of the adrenal gland. The adrenal precursors produced proximal to the enzymatic defect accumulate and promote the synthesis of DHEA and androstenedione through the androgen synthetic pathway. These are converted in the periphery to more potent androgens, resulting in symptoms of hyperandrogenism. Approximately 50% of patients diagnosed with nonclassic congenital adrenal hyperplasia are compound heterozygotes with one of the mutant alleles coding for a severe defect. Typically, patients who are carriers have a normal phenotype. Therefore, most patients with nonclassic adrenal hyperplasia do not demonstrate deficient cortisol production or excess ACTH. It is suggested that most of the androgen excess in nonclassic adrenal hyperplasia arises as a consequence of subtle alterations in enzyme kinetics. Furthermore, some studies report a generalized adrenocortical hyperactivity rather than deficient enzyme activity.

The diagnosis can be established by measuring early morning basal 17-hydroxyprogesterone levels. Levels greater than 800 ng/dL (24.24 pmol/L) are diagnostic of 21-hydroxylase deficiency. However, the elevation in 17-hydroxyprogesterone is often not impressive and does not differ from that observed in PCOS. If the basal 17-hydroxyprogesterone levels are greater than 200 ng/dL (6.06 pmol/L) and less than 800 ng/dL (24.2 pmol/L), a provocative test with ACTH (250 μg intravenously) should be performed. If 17-hydroxyprogesterone levels are greater than 1000 ng/dL (30.30 pmol/L) 1 hour after administration of ACTH, the diagnosis of 21-hydroxylase deficiency can be made. Elimination of a false elevation in 17-hydroxyprogesterone due to ovulation must be excluded by simultaneous measurement of progesterone. A 17-hydroxyprogesterone screening level greater than 200 ng/dL has 100% sensitivity, but only 7% positive predictive value for the diagnosis of nonclassic adrenal hyperplasia. The other rare enzymatic defects that result in nonclassic adrenal hyperplasia can be assessed with measurements of steroid products proximal to the blockade following provocative testing.

Treatment of nonclassic adrenal hyperplasia is similar to that of PCOS, raising a question about whether an etiologic diagnosis is necessary. It is certainly an expense, leading to no change in management except with regard to infertility treatment and management of subsequent pregnancies (see later). Establishing a specific diagnosis may be advisable to facilitate genetic counseling in the woman with adrenal hyperplasia who intends future childbearing and to prepare for in utero treatment should an affected fetus be identified by amniocentesis. Although some experts suggest dexamethasone treatment for symptoms of hyperandrogenism, studies show inconsistent results, and there is concern about the consequent adrenal suppression.

C. Cushing syndrome Chronic glucocorticoid excess, whatever its cause, leads to the constellation of symptoms and physical features known as Cushing syndrome. The most common cause is iatrogenic as a result of glucocorticoid treatment. However, an ACTH-secreting microadenoma (Cushing disease) accounts for more than 70% of cases of endogenous hypercortisolism. Less common causes include primary adrenal disease (tumors or hyperplasia) and ectopic (not hypothalamic-pituitary) ACTH-producing or CRH-producing tumors. Patients with Cushing syndrome

have a range of clinical manifestations that vary with age at onset and etiology. This section will discuss the adult clinical presentation briefly. For a more detailed discussion, see Chapter 9.

Cushing syndrome (noniatrogenic) is rare and occurs in approximately 2.6 patients per million individuals in the population. It is responsible for less than 1% of those individuals who present with hirsutism. Although Cushing syndrome is uncommon, it presents in a manner similar to PCOS and congenital adrenal hyperplasia and needs to be considered in the differential diagnosis of hyperandrogenism and anovulation. Patients with ACTH excess typically have additional clinical features suggestive of glucocorticoid or mineralocorticoid hypersecretion. The most common features include obesity with increased centripetal fat, moon facies, muscle weakness, and striae. Other manifestations may include diabetes, hypertension, and osteoporosis. Women with primary tumors, most commonly adrenocortical carcinomas, tend to have a rapid onset of symptoms and often manifest with severe hyperandrogenism (frank virilization) that includes male pattern baldness, deepening voice, clitoromegaly, and defeminization.

Hirsutism or acne is present in about 60% to 70% of women with Cushing syndrome. However, the exact mechanism of hyperandrogenic effects is not completely known. It is evident that excess ACTH causes hyperstimulation of the zona fasciculata and zona reticularis and results in hypersecretion of cortisol and androgens. It is also known that adrenal tumors may selectively overproduce androgens.

Menstrual irregularities occur in more than 80% of patients with Cushing syndrome. The exact cause of anovulation is unclear. It has already been observed that hyperandrogenemia may have a significant impact on ovulation. However, several studies have shown that glucocorticoids can also suppress the hypothalamic-pituitary axis. Thus, the elevated glucocorticoids may be an additional factor in the pathophysiology of anovulation associated with this syndrome (see Chapters 4 and 9 for the diagnosis and treatment of Cushing syndrome).

D. Androgen-secreting tumor If there is a rapid onset of androgenic symptoms, an androgen-secreting adrenal tumor should be suspected. Elevated testosterone (>200 ng/dL; 6.9 nmol/L) and DHEAS (>700 ng/mL; 19 μmol/L) levels should raise the suspicion of a tumor. However, more than 50% of adrenal androgen-secreting tumors have testosterone levels below 200 ng/dL (6.9 nmol/L). Furthermore, the majority of patients with high testosterone levels do not have tumors. Testosterone levels above 250 ng/dL are only 10% predictive for androgen-secreting tumor. Measurement of DHEAS levels in these patients yields similar inconsistencies. This suggests that laboratory tests have limited value in screening for androgen-secreting tumors, and a clinical history and physical examination are better predictors. The presence of systemic symptoms such as weight loss, anorexia, bloating, and back pain favor an androgen-secreting tumor. If suspicion is high, an abdominal computed tomography (CT) scan confirms the diagnosis. The treatment involves surgical resection, mitotane (adrenolytic), and steroid synthesis inhibitors.

Androgen-secreting tumors can also originate from the ovary. The incidence approximates 1:500 to 1:1000 hyperandrogenic patients. Testosterone levels more than 200 ng/dL (6.9 nmol/L) arouse suspicion, although in 20% of patients with ovarian androgen-producing tumors, testosterone levels are below this value. Again, the best screening procedures are the clinical history and physical examination. In the absence of Cushingoid features, adrenal and ovarian tumors present similarly. Ovarian tumors often have unilateral ovarian enlargement that can be palpated on pelvic examination. Ultrasonography often confirms the diagnosis. In selected cases, selective venous sampling may be performed if CT or sonography cannot identify the source of androgen production.

OBESITY

Obesity is the most prevalent chronic disease in the United States. Adverse health events are dramatically increased in obese subjects (see Chapter 20). These include cardiovascular disease, diabetes, joint disease, respiratory dysfunction, and colon, endometrial, and ovarian cancers. Obesity is often also associated with menstrual irregularities, and the relationship between the two is strengthened the earlier the onset of the obesity. Alterations in sex steroid metabolism are clearly evident in obese females, and the consequences are attenuated release of gonadotropins, which manifests as anovulation. Obese patients have increased production rates and MCRs of androgens; therefore, their serum levels of androgens are largely within the normal range. However, free-testosterone levels tend to be in the high-normal range due to the decreased SHBG levels. A portion of androgen metabolism occurs in body fat. The excess adipose tissue aromatizes these androgens and increases the amount of circulating estrone causing a state of functional hyperestrogenism. In fact, studies have shown that the rate of peripheral conversion of androstenedione to estrone is correlated with body weight. Other studies have illustrated that the conversion of estrone to estradiol in adipose tissue is higher in visceral fat than subcutaneous fat. The increased visceral fat associated with obesity is also associated with hyperinsulinemia, which may have independent effects on ovarian function.

Management of Obesity

The management of obesity includes diet and exercise (see Chapter 20). In patients who have irregular uterine bleeding, an endometrial biopsy must be performed to rule out endometrial cancer. Oral contraceptives are effective at treating the irregular bleeding and reduce the incidence of endometrial cancer. Weight loss reestablishes normal menstrual cycles in the majority of these women.

Anovulation Unrelated to Excess Sex Steroid Production

A. Adrenal insufficiency   Adrenocortical insufficiency may be categorized as primary or secondary. Primary adrenal insufficiency (Addison disease) is caused by destruction of adrenal cortical tissue. Secondary adrenal insufficiency is due to defects in the hypothalamic-pituitary axis resulting in a deficiency in ACTH. Both types lead to cortisol deficiency, which is life threatening.

The main cause of primary adrenal failure is autoimmune destruction of the adrenal cortex. In fact, in the industrialized world, an autoimmune pathogenesis accounts for more than 60% of cases of primary adrenocortical deficiency. The symptoms are typically those of chronic insufficiency and include weakness, fatigue, menstrual disturbances, and gastrointestinal symptoms such as nausea, abdominal pain, and diarrhea. Additional signs may include weight loss, hypotension, and hyperpigmentation of the skin and mucous membranes. These symptoms may appear insidiously with a mean duration of approximately 3 years. Symptoms usually wax and wane until there is complete decompensation.

Autoimmune adrenal failure may occur as an isolated event, but estimates link more than 70% of cases to an APS of two subtypes (types 1 and 2). APS-1 usually presents in childhood with hypoparathyroidism, chronic mucocutaneous candidiasis, adrenal insufficiency, and other autoimmune disorders such as celiac disease and ovarian failure. It is caused by mutations in the autoimmune regulator gene (*AIRE*) and has an autosomal recessive inheritance. APS-2 (Schmidt syndrome) has its onset in adulthood (usually the third decade). Common manifestations include type 1 diabetes mellitus, myasthenia gravis, Hashimoto thyroiditis, ovarian failure, and adrenal insufficiency. Susceptibility to this disorder seems to be inherited as a dominant trait in linkage disequilibrium with the HLA-B region of chromosome 6. Autoimmune adrenal insufficiency is associated with autoantibodies that are directed toward enzymes involved with steroidogenesis. Several studies suggest that the antigens in this disorder include 17α-hydroxylase, 21-hydroxylase, and P450scc. Detection of antibodies directed at these enzymes is helpful in making the diagnosis of autoimmune adrenal failure (see Chapters 2 and 9).

Worldwide, infection—especially tuberculosis—is the most common cause of primary adrenal failure. The adrenal cortex and medulla are involved and may be completely replaced by caseating granulomas. This phenomenon is always associated with other evidence of a tuberculosis infection. Fungal, viral, and bacterial pathogens are less common causes.

The most common cause of secondary adrenal insufficiency is adrenal suppression after exogenous glucocorticoid administration. Adrenal insufficiency is also seen following treatment of Cushing disease and with a variety of hypothalamic-pituitary lesions that result in hypopituitarism. These patients more commonly present with symptoms suggestive of acute adrenal insufficiency. The clinical features include abdominal pain, hypotension, fever, severe volume depletion, and possibly profound shock.

Menstrual disturbances are a frequent presentation in patients with adrenal insufficiency. Autoimmune adrenal insufficiency is often accompanied by gonadal failure. In the two forms of APS, POF develops in APS-1 in 50% of patients and in APS-2 in 10% of patients. It has been shown that antibodies—particularly to CYP17α and P450scc—are associated with POF. Other causes of adrenal failure are associated with menstrual disorders more than 25% of the time. The cause of anovulation is not known with certainty, but chronic illness itself is probably responsible.

Diagnosis and treatment are discussed in Chapter 9. The screening process includes blood chemistries and basal cortisol levels. The diagnosis is confirmed with provocative tests using exogenous ACTH.

B. Thyroid disorders   The prevalence of overt thyroid dysfunction is 1% to 2% in women of reproductive age. Thyroid disorders can develop secondary to an insult in the hypothalamus, pituitary, or thyroid, the latter being most common. In order to understand the pathophysiology of these disorders, it is important to be familiar with the normal physiologic regulation (see Chapter 7). This section will briefly review the common causes of hyperthyroidism and hypothyroidism, their manifestations after the onset of puberty, and their impact on reproductive function.

1. **Hyperthyroidism**—Hyperthyroidism is the clinical syndrome associated with excessive thyroid hormone activity. The clinical presentation of hyperthyroidism (thyrotoxicosis) depends on the age at onset and the degree of thyrotoxicosis. The clinical manifestations can involve most organ systems, and the presentation ranges from asymptomatic to thyroid storm. Typical features include nervousness, malaise, palpitations, heat intolerance, weight loss, and inability to concentrate. Additional features may involve the eyes and include lid lag, proptosis, and ophthalmoplegia. The reproductive abnormalities include menstrual abnormalities, infertility, and spontaneous abortions.

   The most common cause of hyperthyroidism is an autoimmune process that affects thyroid hormone production (Graves disease). Antibodies bind to the TSH receptor and stimulate the thyroid gland to secrete increased amounts of thyroid hormone. Less common causes include subacute thyroiditis, toxic multinodular goiter, and struma ovarii.

   Excess thyroid hormone has an impact on sex steroids. Thyroid hormones stimulate hepatic production of SHBG. As a result, total serum estradiol, estrone, testosterone, and dihydrotestosterone are increased, yet free levels of these hormones remain within the normal range. The metabolic clearance pathways appear to be altered, which can be explained in part by the increased binding. The conversion rates of androstenedione to estrogen and testosterone are increased. The significance of the alterations in metabolism has not been determined.

   Menstrual irregularities frequently occur in hyperthyroid states. The exact mechanism is unclear. Altered levels of TRH and TSH do not appear to have a significant impact on the HPO axis. The LH surge may be impaired, although studies have shown that hyperthyroid patients have normal FSH and LH responses to exogenous GnRH. It is possible that the weight loss and psychologic disturbances associated with this disease may contribute to the menstrual abnormalities. It is interesting to note that endometrial biopsies of many amenorrheic patients with hyperthyroidism have demonstrated a secretory endometrium, indicating that many of these women remain ovulatory. Menstrual abnormalities return to normal with treatment.

2. **Hypothyroidism**—A more common disorder than hyperthyroidism in women of reproductive age, hypothyroidism results from inadequate thyroid hormone production. Manifestations can involve almost any organ system, and the presentation can range from asymptomatic to myxedema coma. Common symptoms include lethargy, memory defects, cold intolerance, dry skin, hair loss or occasionally excess hair growth, deepening of the voice, nausea, and constipation. Physical findings include somnolence; bradycardia; mild hypertension; dry skin; periorbital puffiness; nonpitting edema of the hands, face, and ankles; and decreased tendon reflexes. Reproductive abnormalities include menstrual disorders, infertility, and spontaneous abortions.

   The most common cause is autoimmune destruction of the thyroid (Hashimoto thyroiditis). It is mediated by humoral and cell-mediated processes. The antibodies are directed toward thyroglobulin (anti-Tg) and thyroid peroxidase (anti-TPO antibody). Histologic specimens show lymphocytic infiltration. Other causes include ablative therapy of the thyroid gland (postsurgery or radioactive iodine), end-stage Graves disease, and transient thyroiditis (viral, drug-induced, or postpartum).

   Inadequate thyroid hormone levels influence the metabolism of sex steroids. The production of SHBG is decreased. As a result, serum estradiol and testosterone concentrations are decreased, but free-hormone levels remain within the normal range. However, the metabolism of these steroids is altered and differs from that found in individuals with hyperthyroidism.

   The mechanism underlying menstrual abnormalities in hypothyroidism is incompletely understood. Because primary hypothyroidism is associated with elevated serum PRL in up to one-third of patients, it is plausible that hyperprolactinemia is a contributing factor (see Chapter 4). However, menstrual abnormalities are also observed in the absence of elevated PRL. Alterations in FSH and LH levels have been investigated, and studies have shown inconclusive results, although several studies suggest that the midcycle surge is absent. Menstrual function returns to normal with thyroid hormone replacement.

   Thyroid dysfunction often presents with nonspecific symptoms, which delays the diagnosis. If only menstrual abnormalities are present, it is prudent to screen for thyroid abnormalities. Most cases are detected by TSH assays. Confirmation is obtained with a repeat TSH level and serum thyroid hormone levels.

OUTFLOW TRACT DISORDERS

The true prevalence of Müllerian tract abnormalities is not known. It is reported in as many as 45% of women. The reproductive consequences depend on the type of abnormality identified. A septate uterus is the most common defect described. These patients often present with infertility or obstetric complications. Other abnormalities include unicornous, bicornous, and didelphic uteri. These abnormalities present most commonly with reproductive or obstetric complications. Müllerian agenesis, androgen insensitivity syndrome, and congenital outflow obstruction defects are abnormalities that present with primary amenorrhea with no history of menses. In the following section these two abnormalities are discussed in greater detail.

A. Müllerian agenesis   Müllerian agenesis (Mayer-Rokitansky-Küster-Hauser syndrome) is the second most common cause of primary amenorrhea. It is a congenital condition that occurs in

1 in 5000 female births. These individuals have normal ovarian development, normal endocrine function, and normal female sexual development. The physical findings are a shortened or absent vagina in addition to absence of the uterus, although small masses resembling a rudimentary uterus may be noted (see the section "Embryology and Anatomy" at the beginning of this chapter). About one-third of patients have renal abnormalities, and several have had bone abnormalities and eighth nerve deafness. These individuals have a 46,XX karyotype.

The exact cause has not been identified. It is known that regression of Müllerian structures in males is controlled by AMH, which is secreted by the Sertoli cells of the testis. One hypothesis assigns the underlying defect to an activating mutation of either the *AMH* gene or its receptor. The genes for both AMH and AMH receptor have been investigated, but no mutations have yet been identified in patients with this syndrome.

B. Androgen insensitivity syndrome   Androgen insensitivity syndrome (AIS) presents in somewhat the same way as Müllerian agenesis. The presentation differs in that individuals with complete AIS have minimal sexual hair. These patients have a male karyotype with a mutation of the androgen receptor on the X chromosome. They have normal testicular development and endocrine function. However, because the internal and external male sexual structures need testosterone for development, they are absent. This results in a female phenotype. Because the testis still secretes AMH, Müllerian regression does occur. Secondary sexual characteristics (female) develop as a result of peripheral conversion of testosterone to estradiol, effectively resulting in unopposed estrogen stimulation.

The diagnosis of either disorder is entertained when pelvic examination reveals a short or absent vagina and no uterus on rectal examination. Confirmation of absent uterus can be obtained with ultrasound, MRI, or laparoscopy. These two disorders can usually be differentiated based on physical examination; patients with AIS have no pubic hair. However, the differential diagnosis becomes more difficult when patients have incomplete AIS. A testosterone level and karyotype can easily differentiate the two syndromes. This disorder is further discussed in Chapter 14.

C. Congenital outflow obstruction   Transvaginal septum and imperforate hymen are typical obstructive abnormalities. These patients usually present with cyclic lower abdominal pain and amenorrhea. The physical findings are a shortened or absent vagina. However, this syndrome differs from Müllerian agenesis in that the pelvic organs are present. Behind either defect is old blood that has not escaped with menses. The differential diagnosis is sometimes difficult, although bulging of the introitus suggests imperforate hymen, because the defect is thinner than a transvaginal septum.

The embryologic formations of the transvaginal septum and imperforate hymen are similar but not identical. Transvaginal septum is due to failure of complete canalization of the vaginal plate (see the section "Embryology and Anatomy" at the beginning of this chapter). The septum can vary in thickness and can be located at any level in the vagina. The hymen represents the junction of the sinovaginal bulbs and urogenital sinus. Typically, the hymen is perforated during fetal development. The hymen is thin and is always at the junction of the vestibule and vagina. It is important to distinguish these defects because the surgical correction procedures are different and require different levels of expertise.

D. Asherman syndrome   Intrauterine adhesions or synechiae (Asherman syndrome) are an acquired condition that may obliterate the endometrial cavity. These patients usually present with a range of menstrual disturbances, infertility, and recurrent spontaneous abortions. The most frequent symptom is amenorrhea.

Intrauterine adhesions result from damage to the endometrial basal layer. A common antecedent factor is a surgical procedure within the uterine cavity, and most often it is endometrial curettage that occurs shortly after pregnancy. The concurrent presence of infection or heavy bleeding increases the risk. Endometrial tuberculosis and septic abortion are rare causes.

The diagnosis is entertained after demonstrating no withdrawal bleeding after administration of estrogen and progesterone. Confirmation is made with a hysterosalpingogram, saline sonogram, or hysteroscopy. Treatment involves lysis of adhesions and hormonal therapy.

MENOPAUSE

The ovary is unique in that the age at which it ceases to function in women appears to have remained constant despite the increase in longevity experienced by women over the last century. Because the loss of ovarian function has a profound impact on the hormonal milieu in women and the subsequent risk of the development of disease resulting from the loss of estrogen production, improving our understanding of reproductive aging is critical for optimal female health.

Human follicles begin development in the fourth gestational month. Approximately 1000 to 2000 germ cells migrate to the gonadal ridge and multiply, reaching a total of 6 to 7 million around the fifth month of intrauterine life. At this point, multiplication stops and follicle loss begins, declining to approximately 1 million by birth. In the human male, the germ cells become quiescent and maintain their stem cell identity. In contrast, in the human female, between weeks 12 and 18, the germ cells enter meiosis and differentiate. Thus, in the female, all germ stem cells have differentiated prior to birth. In the adult woman, the germ cells may remain quiescent, may be recruited for further development and ovulation, or may be destroyed by apoptosis. Over time, the population of oocytes is depleted (without regeneration) through recruitment and apoptosis until less than a thousand oocytes remain and menopause ensues. Approximately 90% of women experience menopause at a mean age of 51.2 years (range, 46-55). The remainder experiences menopause prior to age 46 (often termed early menopause), with 1% of women experiencing menopause before age 40 years (POF, as discussed earlier).

Understanding ovarian aging has been difficult. The variability in definitions has made comparisons from study to study difficult.

Menarche					**FMP (0)**				
Stage	−5	−4	−3b	−3a	−2	−1	+1a	+1b	+1c	+2
Terminology	**Reproductive**				**Menopausal transition**		**Postmenopause**			
	Early	Peak	Late		Early	Late	Early			Late
					Perimenopause					
Duration	*Variable*				*Variable*	1-3 years	2 years (1+1)		3-6 years	*Remaining lifespan*
Principal criteria										
Menstrual cycle	Variable to regular	Regular	Regular	Subtle changes in flow/length	*Variable length* persistent ≥7-day difference in length of consecutive cycles	Interval of amenorrhea of ≥60 days				
Supportive criteria										
Endocrine FSH AMH Inhibin B			Low Low	Variable* Low Low	↑ Variable* Low Low	↑ >25 IU/L** Low Low	↑ Variable Low Low	Stabilizes Very Low Very Low		
Antral follicle count			Low	Low	Low	Low	Very Low	Very Low		
Descriptive characteristics										
Symptoms						Vasomotor symptoms *likely*	Vasomotor symptoms *most likely*			*Increasing* symptoms of urogenital atrophy

* Blood draw on cycle days 2-5 ↑ = elevated
**Approximate expected level based on assays using current international pituitary standard

FIGURE 13–16 The stages of Reproductive Aging Workshop +10 staging system for reproductive aging in women. (Reproduced with permission from Harlow SD, Gass M, Hall JE, et al. Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. *Menopause*. 2012 Apr;19(4):387-395.)

The participants in the Stages of Reproductive Aging Workshop (STRAW) +10 developed criteria for staging female reproductive aging (Figure 13–16). They utilized menstrual cyclicity and early follicular FSH levels as the primary determinants for this staging system. Stages −5 to −3a include the reproductive interval; stages −2 to −1 are termed the menopausal transition; and stages +1a through +2 are the postmenopause (see Figure 13–16). The telltale sign of the early menopausal transition is increased variability in menstrual cyclicity (>7 days) that is persistent in women with elevated FSH levels, low AMH, and low antral follicle count. Persistence is defined as recurrence within 10 cycles of the first variable cycle length. The duration of this transition is variable among women. The late menopausal transition is believed to last 1 to 3 years and is associated with vasomotor symptoms and interval of amenorrhea of 60 days or longer. This menopausal transition ends with the final menstrual period, which cannot be recognized until after 12 months of amenorrhea. Early postmenopause is defined as the first 6 years following the final menstrual period. The presence of urogenital atrophy is not thought to occur until the late postmenopause which begins 6 years after the final menstrual period and continues until death.

This system is said to include endocrinologic aspects of ovarian aging, but it still depends largely on menstrual cyclicity as a key indicator of ovarian age. The system includes measurement of FSH and AMH; however, by the time FSH is elevated or AMH is low, even in the face of cyclic menstrual cycles, oocyte depletion has already proceeded to such an extent that fertility (as a marker of reproductive aging) is significantly diminished. Evidence suggests that genetic and environmental factors influence both age at menopause and the decline in fertility, although the specific nature of these relationships is poorly characterized. Premature menopause can be due to failure to attain adequate follicle numbers in utero or to accelerated depletion thereafter. Potentially, either of these causes could be affected by genetic and environmental factors. The timing of menopause has a consistent impact on overall health with respect to osteoporosis, cardiovascular disease, and cancer risk. Over the next decade, it is estimated that more than 40 million women in the United States will enter menopause.

OOCYTE DEPLETION

As discussed earlier, the leading theory regarding the onset of menopause relates to a critical threshold in oocyte number. The theory that menopause is primarily triggered by ovarian aging is supported by the coincident occurrence of follicular depletion, elevation of gonadotropins, and menstrual irregularity with ultimate cessation.
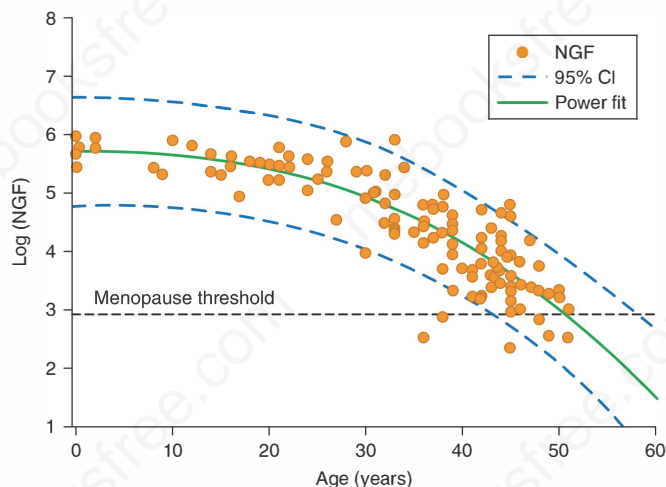
FIGURE 13–17 A new model of reproductive aging: the decline in ovarian nongrowing follicle number (primordial follicles) from birth to menopause. (Reproduced with permission from Hansen KR et al. A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. *Hum Reprod.* 2008 Mar;23(3):699-708.)

In 1992, Faddy and colleagues developed a mathematical model to predict the rate of follicular decline. They utilized data from multiple sources to construct a model that ultimately showed a biexponential decline with an acceleration in oocyte loss beginning in the late thirties. More recently, Hansen et al utilized newer technology and a single population of specimens. They fit the data to multiple models and found that a pattern of gradually increasing atresia fit the decline in nongrowing follicles (Figure 13–17).
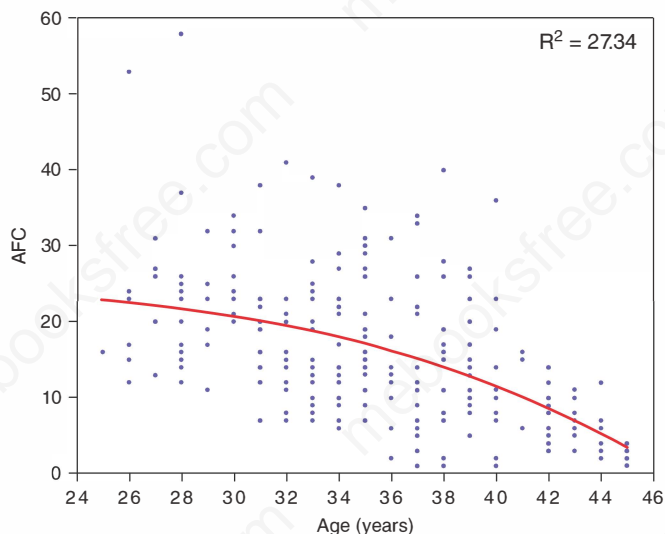
FIGURE 13–18 The antral follicle count (AFC) which declines with age in a Caucasian population is best described as a gradual acceleration in decline with age. Relationship of total AFC and chronological age. (Reproduced with permission from Rosen M et al. Antral follicle count: absence of significant midlife decline. *Fertil Steril.* 2010 Nov;94(6):2182-2185.)

This same model most closely follows the decline in the antral follicles—small (2-10 mm) growing follicles seen on ultrasound—seen across the reproductive-age span (Figure 13–18). As noted earlier, the loss of follicles (oocytes) is a constant process beginning in utero and continuing throughout reproductive life, even during pregnancy and in the absence of ovulation, until approximately 1000 follicles/oocytes remain and menopause ensues.

ENDOCRINE SYSTEM CHANGES WITH AGING

The entire endocrine system changes with advancing age. The somatotrophic axis begins to decline in the fourth decade, prior to the decline in ovarian function. This decline is accelerated in the face of ovarian failure and may act to accelerate the decline in ovarian function. However, pituitary concentrations of GH as well as ACTH and TSH remain constant into the ninth decade. While the thyroid gland undergoes progressive fibrosis with age and concentrations of T_3 decline by 25% to 40%, elderly patients still remain euthyroid. Beta cell function also undergoes degeneration with aging such that by age 65 years, 50% of subjects have abnormal glucose tolerance tests. Frank diabetes is less frequent, however, occurring in approximately 7% of the population. The female reproductive system, on the other hand, undergoes complete failure at a relatively early age.

Secretion of reproductive hormones during the menopausal transition (MT) was previously thought to decline progressively in a linear fashion, but hormone levels have since been shown to fluctuate widely. Studies in large cohorts of women have demonstrated that circulating FSH concentrations rise progressively during the MT. The initial monotropic rise in FSH is attributed to a decrease in ovarian inhibin secretion rather than to a decrease in estradiol production.

Inhibin and activin are proteins produced by the granulosa cells and have been shown to play major roles during the MT. Inhibin consists of a covalently bound dimer with an α subunit and one of two different β subunits, designated as β_A and β_B; the resulting heterodimers are known as inhibin A and inhibin B. Inhibin A is secreted by the corpus luteum and inhibin B by antral and dominant follicles. Consequently, inhibin A levels increase during the luteal phase, and inhibin B concentrations rise during the follicular phase. Both inhibins inhibit pituitary FSH secretion. Activins are a related class of proteins that stimulate pituitary FSH release. The activin molecule is a homodimer composed of two covalently linked inhibin β subunits, designated as activin A ($\beta_A\beta_A$) and activin B ($\beta_B\beta_B$).

AMH is also secreted by the granulosa cells of secondary and preantral follicles. Circulating concentrations remain relatively stable across the menstrual cycle and correlate with the number of early antral follicles. Levels of AMH decrease markedly and progressively across the MT to a time point approximately 5 years prior to the final period when levels are below the limit of detection.

During the late reproductive stage (stage –3), follicular phase inhibin B levels decrease as FSH concentrations rise. As the MT progresses, luteal phase inhibin A levels also decline. Activin A concentrations also are elevated in perimenopausal women. Whereas activins clearly play a local role in regulating pituitary FSH secretion, their ability to act as endocrine factors to influence the production of FSH has not been established. Thus, a decrease in secretion of inhibin A and inhibin B, and a corresponding increase in activin production may favor increased FSH secretion in the absence of any decrease (and perhaps an increase) in estradiol production.

Estrogens/Progesterone

The main circulating estrogen during the premenopausal years is 17β-estradiol. Levels of this hormone are controlled by the developing follicle and resultant corpus luteum. The fact that oophorectomy reduces peripheral estradiol levels from 120 to 18 pg/mL confirms that over 85% of circulating estradiol is derived from the ovary. In perimenopausal women, estradiol production fluctuates with FSH levels and can reach higher concentrations than those observed in young women under age 35. Estradiol levels generally do not decrease significantly until late in the MT. Estradiol levels may be quite variable, with chaotic patterns and occasionally very high or very low levels. This dramatic variability may lead to an increase in symptomatology during the perimenopausal years (stages –2 to –1). As peripheral gonadotropins rise, LH pulsatile patterns become abnormal. There is an increase in pulse frequency, with a decrease in GnRH inhibition by opioids. Despite continuing regular cyclic menstruation, progesterone levels during the early MT are lower than in women of mid-reproductive-age and vary inversely with body mass index. Women in the late MT exhibit impaired folliculogenesis and an increasing incidence of anovulation, compared to mid-reproductive-aged women. Essentially all estradiol in postmenopausal women is derived from peripheral conversion from estrone.

The predominant estrogen in the postmenopausal woman is estrone, with a biologic potency approximately one-third that of estradiol. The circulating levels of estrone (and estrone sulfate) in older women are approximately one-third to one-half of the concentration observed in women of reproductive age (Table 13–7).

TABLE 13–7 Plasma concentrations of sex steroids in premenopausal and postmenopausal women.

Steroid (concentration)	Premenopausal	Postmenopausal
Estrone (pg/mL)	50->125[a]	30-35
Estrone sulfate (pg/mL)	1000-1800[a]	350
Estradiol (pg/mL)	50->250[a]	10-20
DHEAS (ng/mL)	1-3	0.1-2.7
Testosterone (ng/mL)	<1	<1

[a]Plasma concentration depends on the phase of menstrual cycle.
Abbreviation: DHEAS, dehydroepiandrosterone sulfate.

This is due to estrone production, resulting largely from peripheral aromatization of androstenedione (see earlier). This aromatase activity increases with aging by two- to fourfold and is further amplified by the increased adiposity that typically accompanies the aging process (see earlier). Estrone and estradiol production rates during the postmenopausal years are 40 and 6 μg/d, respectively. This compares with 80 to 500 μg/d for estradiol during the reproductive years.

Androgens

In contrast to estrogens, circulating androgens (DHEAS, androstenedione, and testosterone) decrease less dramatically after physiologic menopause (see Table 13–7). DHEAS undergoes a linear reduction with aging in both men and women (~2% per year starting at age 30), but there is no specific decline associated with menopause. Changes in DHEAS levels have been associated with alterations in body composition with aging. Androstenedione production similarly decreases with aging, but the circulating levels are affected less because ovarian secretion is maintained, albeit at a reduced rate. Testosterone levels do not vary appreciably during the MT. Testosterone levels decrease after menopause. Historically it was believed the postmenopausal ovary produced a larger percentage of testosterone (50%) than did the premenopausal ovary. However, the literature is conflicting on the origin of androgens in postmenopausal women. The conventional view was that the adrenals, ovaries, and periphery all contributed to circulating androgen levels in postmenopausal women. Contemporary data suggest that androgen production after menopause is largely derived from adrenal precursors. This is supported by studies that have shown that postmenopausal women receiving dexamethasone suppression, or who have endogenous adrenal insufficiency, have undetectable levels of circulating androgens. Investigators have also demonstrated that the postmenopausal ovary has no appreciable enzymatic activities capable of generating sex steroids. The fact that women of reproductive age who undergo a bilateral oophorectomy have less testosterone and androstenedione than menopausal women with intact ovaries challenges this concept. However, the conflict may be explained by the hypothesis that the postmenopausal ovary produces androgens for a limited time.

Hypothalamic/Pituitary

Significant additional information regarding the hormonal changes during the menopausal transition is being developed in the multiethnic, community-based, longitudinal study of perimenopausal women at seven sites throughout the United States, The Study of Women's Health Across the Nation (SWAN). While most attention has been devoted to the role of the ovarian and oocyte decline in the onset of menopause, new evidence suggests potential alterations in the hypothalamic-pituitary feedback system. In evaluating the ability of estradiol to stimulate an effective LH surge, data from SWAN suggest the MT is characterized by three distinct hormonal patterns: (1) estrogen rise with rise in LH

but anovulation; (2) estrogen rise without a concomitant rise in LH; or (3) failure of rise in either estradiol or FSH. Additionally, there may be ethnic differences in the sensitivity of the pituitary to negative feedback.

MENOPAUSAL CONSEQUENCES

Given the endocrinologic changes associated with aging, many symptoms appearing in the aging female may be due to estrogen deficiency or diminished androgen or GH secretion. Disorders that are definitely due to estrogen deprivation include vasomotor symptoms and urogenital atrophy. Osteoporosis is also thought to be due largely to estrogen deficiency, and it may be exacerbated by the relative decline in GH levels. The same may be said for the hormone-related increase in the prevalence of atherosclerotic cardiovascular disease and psychosocial symptoms, including insomnia, fatigue, short-term memory changes, and possibly depression. Both DHEAS and GH may impact these phenomena as well. Most women who are symptomatic during the MT present with frequent or excessive bleeding or with hot flashes and other symptoms of estrogen deficiency. Other common symptoms during the MT include decreased libido, forgetfulness, vaginal dryness, and urinary incontinence.

Mood disorders also are increased during the MT. Community-based surveys have shown that perimenopausal women report significantly more psychological distress and have an increased risk for significant depression, compared with premenopausal or postmenopausal women. In an 8-year longitudinal study of premenopausal women having no prior history of depression, a depressive disorder was more likely to be diagnosed during the MT than in the premenopausal years (OR, 2.5; 95% CI, 1.25-5.02, $p = 0.01$).

Vasomotor Symptoms

Vasomotor symptoms (hot flushes) are experienced with greatest frequency during stages −1 and +1, with about 75% to 85% of women complaining of this symptom. Although 80% of the US women have symptoms lasting for at least 1 year, only 25% of women are still symptomatic at 5 years after the final menstrual period. Studies of hot flushes with external monitoring of skin temperature and resistance have shown a frequency of approximately 54 ± 10 minutes. In sleep studies, hot flush frequency has been shown to interrupt rapid eye movement sleep and may contribute to some of the psychosocial complaints. Hot flushes are temporally correlated with pulses of LH, but exogenous LH does not induce a flush, suggesting that there is some central mediator leading to both the flush and the elevation in LH.

Sleep disturbances, most likely the result of vasomotor activity, also are very common during the MT. In longitudinal studies of perimenopausal women, the prevalence of sleep disturbances has ranged from 32% to 40% in the early MT and from 38% to 46% in the late MT.

Vasomotor symptoms during the MT can be treated with hormone therapy (HT) using estrogen or progestin alone or in combination, neuroactive agents, or other nonhormonal alternatives.

Estrogen therapy provides the best treatment for severe vasomotor symptoms, reducing their frequency and severity in 80% to 85% of women over 12 weeks compared with 30% of women receiving placebo. Hormone therapy also may help in the management of depression associated with the MT. In a randomized, placebo-controlled, 12-week trial involving perimenopausal women ages 40 to 55, symptoms of depression were improved in 68% of women receiving unopposed estrogen treatment (0.1 mg transdermal estradiol patch), compared with 20% of those receiving placebo ($p = 0.001$). In general, HT should be used at the lowest doses to manage menopausal symptoms. Testing estrogen levels is not necessary and should not guide treatment.

Other drugs have been tried in women for whom estrogen is contraindicated, although none have the efficacy of HT for the treatment of menopausal symptoms. These alternatives include transdermal clonidine, ergot alkaloids, and, more recently, selective serotonin reuptake inhibitors and gabapentin. High-dose progestins may also produce some relief.

Genital Atrophy

The vagina, vulva, urethra, and bladder trigone not only share embryonic proximity but all contain estrogen receptors. Atrophy begins in stage −2 to −1. The most common symptoms include itching and vaginal thinning, with decreased distensibility and reduced secretions, leading to vaginal dryness and pain with intercourse. This and the change in pH with resultant changes in vaginal flora increase the incidence of vaginal and urinary tract infections. Estrogen is the treatment of choice, and treatment must continue for at least 1 to 3 months for symptomatic improvement to be noted. The systemic dosage necessary for vaginal protection is somewhat higher than that needed for bone protection (see later), and local therapy by means of creams or vaginal rings may thus be advisable to limit systemic absorption. It should be noted, however, that vaginal absorption of steroids is quite efficient once estrogenization and revascularization have occurred. If the goal is to limit systemic absorption, slow-release rings may be superior to estrogen creams.

Vaginal estrogen frequently improves symptoms of urinary frequency, dysuria, urgency, and postvoid dribbling. Its direct effect to improve stress incontinence is less clear.

Osteoporosis

Osteoporosis is a condition in which bone loss is sufficient to allow fracture with minimal trauma. Major risk factors for the development of osteoporosis are the peak bone density attained in the late teens and early twenties (stressing the importance of bone building in the young) and the rate of loss (accelerated with estrogen deficiency). Primary or *senile* osteoporosis usually affects women between the ages of 55 and 70 years. The most common sites include the vertebrae and the long bones of the arms and legs. Secondary osteoporosis is caused by a specific disease (such as hyperparathyroidism) or medication usage (such as glucocorticoids) (see Chapter 8).

Menopausal bone loss begins before the final menstrual period during stage −1. Postmenopausal osteoporosis causes over 1.3 million

fractures annually in the United States. Most of the more than 250,000 hip fractures are due to primary osteoporosis, and—given that 15% of patients die within a year after a hip fracture and 75% of patients lose their independence—the social costs, not to mention the financial costs, are great.

Bone loss following natural menopause is approximately 1% to 2% per year compared with 3.9% per year following oophorectomy. A woman's genetic background, lifestyle, dietary habits, and coexisting disease also influence the development of osteoporosis. Cigarette smoking, caffeine usage, and alcohol consumption also negatively affect bone loss, whereas weight-bearing activity appears to have a positive influence. See Chapter 8 for a detailed discussion of bone and mineral metabolism and osteoporosis.

A. Osteoporosis treatment

1. **Estrogen therapy**—Estrogen therapy acts via the inhibition of bone resorption. Both BMD and fracture rate are improved with estrogen therapy. However, with cessation of estrogen therapy, there is a rapid and progressive loss of bone mineral content. By 4 years after therapy, bone density is no different from that of patients who were never treated with estrogen. Estrogen is approved for prevention of osteoporosis, and there is also some support for its usage as a treatment modality in established disease. Dosages of 0.625 mg of conjugated estrogens orally daily—and, more recently, as low as 0.3 mg—have been shown to slow bone loss and provide adequate protection against the development of osteoporosis. Higher dosages may be required to treat existing disease.

2. **Alternative therapies for osteoporosis**

    a. **Calcitonin**—Calcitonin is a hormone normally secreted by the thyroid gland. Calcitonin (salmon) is available as a nasal spray specifically developed to decrease local side effects caused by subcutaneous injection. Although few studies have been performed and no data are available regarding reduction in hip fracture, it does seem to be especially beneficial for women with a recent and still painful vertebral fracture. Intranasal calcitonin has also been shown to improve spinal bone density and decrease the vertebral fracture rate in established osteoporosis. The increase in bone density appears to peak in as little as 12 to 18 months. The waning effects of calcitonin therapy over time may be due to downregulation of calcitonin receptors on osteoclasts and/or the development of neutralizing antibodies.

    b. **Bisphosphonates**—These compounds are analogs of pyrophosphates and have a high affinity for hydroxyapatite in bone matrix. The basic structure of bisphosphonates allows a large number of manipulations of the basic molecule, producing different types of bisphosphonates that vary considerably in their potency on bone. In order of increasing potency are pamidronate, alendronate, risedronate, ibandronate, and zoledronic acid.

    Alendronate has been evaluated extensively in patients with osteoporosis. Alendronate has been shown to inhibit markers of bone remodeling and increase BMD at the lumbar spine, hip, and total body and reduces fracture risk. Alendronate is taken orally; the recommended doses are 70 mg once weekly or 10 mg daily; alendronate must be taken according to a strict dosing schedule (in the morning on an empty stomach, with the patient required to remain upright for 30 minutes thereafter). The medication has very poor bioavailability (approximately 1%), and for that reason these instructions must be meticulously obeyed. Alendronate also has a propensity for causing irritation of the esophagus and stomach, especially in women with preexisting esophageal reflux, gastric or duodenal disease. Risedronate is similarly effective in the dosage of 35 mg weekly or 5 mg once daily, and the same dosing regimen is recommended. Ibandronate has the easiest schedule for administration 150 mg once monthly.

    Increases in bone density with alendronate, risedronate, and ibandronate are greater than what is seen with calcitonin and similar to what is seen with HT. The escape phenomenon seen with calcitonin is not seen with these oral bisphosphonates.

    An important issue concerning long-term administration of bisphosphonates relates to their long half-lives in bone and their incorporation into the bone matrix and the potential for rare adverse events such as atypical fractures and osteonecrosis of the jaw.

    c. **SERMs**—Raloxifene is the first selective estrogen receptor modulator (SERM) approved for prevention and treatment of osteoporosis and for the prevention of breast cancer. SERMs act selectively as estrogen receptor agonists in some tissues (bone and heart) and antagonists in others (breast and uterus and possibly brain). Data with raloxifene show preservation of BMD, albeit less well than that seen with alendronate, risedronate, ibandronate, or HT, and spinal fracture data support a protective effect.

    It is believed that the differential effects of estrogens and antiestrogens are related to the transcriptional activation of specific estrogen response elements. Two different domains of the estrogen receptor (AF-1 and AF-2) are responsible for this transcriptional activation. Estrogens and antiestrogens appear to act via different domains, leading to their differential effects. Both act to maintain BMD—at least partially—via regulation of the gene for TGF-β.

    Recently a tissue selective estrogen complex (TSEC), a combination of conjugated estrogen plus the SERM bazodoxifene, was released for the prevention of osteoporosis. This agent is designed to preserve BMD while controlling the side effects commonly seen with SERM therapy (eg, hot flashes) by estrogen.

    d. **Calcium and vitamin D**—These are critical adjuvants for any type of antiresorptive therapy. Decreased ability to absorb calcium among older women is due in part to impaired vitamin D activation and effect. Older women may have limited exposure to sunlight, and their dietary vitamin D intake may be lower than that of younger women.

    Daily intakes of 1000 to 1200 mg of calcium and 800 to 1200 IU of vitamin D are probably sufficient to reduce the risk of fragility fractures very modestly (~10%).

    e. **Anabolic therapy**—The only anabolic therapy currently available for the treatment of severe osteoporosis parathyroid hormone (1-34) or teriparatide was approved for clinical use in 2001. Details about its efficacy and use are described in Chapter 8.

f. Antibody to receptor activator of nuclear factor kappa B ligand (RANK-L) therapy—This agent is a neutralizing monoclonal antibody to RANK-L that blocks osteoclastogenesis and is highly effective in reducing spine, hip, and nonvertebral fractures. Its mechanism of action and efficacy are discussed in Chapter 8.

Atherosclerotic Cardiovascular Disease

Cardiovascular disease is the number one cause of mortality in both men and women in Western society. This is largely attributed to age and lifestyle. Lifestyle modifications are known to decrease the incidence of atherosclerotic cardiovascular disease. For women, cardiovascular disease is largely a disease of the postmenopause. Women now spend more than a third of their lives in the postmenopausal years, and preventive measures are thus of paramount importance. Although a large body of observational evidence supported a protective effect of ET on cardiovascular disease, observational data are limited by the confounding variables of patient self-selection. Animal and *in vitro* studies, as well as assessment of surrogate markers in women, also showed a positive effect of ET and HT against cardiovascular disease development. However, several large randomized, controlled studies fail to support a protective role for HT in preventing cardiovascular events.

One of the first such trials was the Heart and Estrogen-Progestin Replacement Study (HERS), a secondary prevention trial that evaluated the use of daily HT (0.625 mg conjugated estrogens plus 2.5 mg medroxyprogesterone acetate [MPA]) in 2763 postmenopausal women with a mean age of 66.7 years and preexisting vascular disease. The study failed to demonstrate any overall difference in vascular events. This occurred despite improvements in lipid parameters in those patients receiving HT. The Estrogen Replacement and Atherosclerosis (ERA) Trial compared 3.2 years of treatment with estrogen, combined estrogen and progestin, and placebo in postmenopausal women aged 42 to 80 years. It also failed to demonstrate a significant difference in the rate of progression of coronary atherosclerosis between the three groups. The importance of this study was the inclusion of an estrogen-only arm.

The Women's Health Initiative (WHI) was the first large randomized study to look at primary prevention of cardiovascular disease. This study compared (1) the combination of conjugated equine estrogen (CEE) and MPA with placebo and (2) CEE to placebo. It was designed to assess the overall risks and benefits of HT in a prospective randomized fashion.

The WHI demonstrated that there was an unacceptable risk profile for the combination HT arm of the trial. There was an increase in the incidence in breast cancer (an increase of 8 cases per 10,000 women) with no cardiovascular protection (and potentially increased cardiovascular risk). There was, in fact, an increase in venous thromboembolism, strokes, and coronary heart disease. The risk of stroke and thromboembolism continued for the 5 years of study, whereas most of the coronary heart disease was limited to the first year of treatment. There were, however, documented decreases in the risk of fracture and colon cancer.

In the CEE-only arm, there was an increased risk of stroke and a decreased risk in hip fractures compared to placebo. This study showed that the use of ET had no protective effect on coronary heart disease. Interestingly, there appeared to be a trend toward a reduction in breast cancer (0.77, CI 0.59-1.01) with use of ET. More recent studies and reevaluation of large datasets have suggested that combined estrogen/progestin has a more profound increased risk on breast cancer, than estrogen alone or cyclic progestins, consistent with increase in mitogenic activity of the breast during the normal luteal phase when progesterone levels are high.

Treatment—Summary

As noted earlier, long-term use of HT in older menopausal women has been associated with increased risks for venous thromboembolism, coronary events, stroke, and breast cancer. Although short-term treatment of symptomatic women during the MT likely poses significantly fewer risks, HT generally should be used in the lowest effective dose and for the shortest time required. Testing estrogen levels is not necessary and should not guide treatment. Data suggest the use of HT does not increase the risk of breast cancer, on a yearly basis, more than would continuing spontaneous cycles for any given age. This suggests treatment of symptomatic women in their forties may not increase breast cancer risk over normal cycling. Low-dose estrogen regimens (conjugated equine estrogens, 0.3 mg daily, or its equivalent) can achieve as much as a 75% reduction in vasomotor symptoms over 12 weeks, approaching the efficacy of standard-dose HT regimens, and may be associated with fewer risks and side effects. The decision to use HT should be made only after first carefully reviewing its risks and benefits for the individual.

The relative safety of HT during the MT has not been thoroughly investigated. The results of one observational study have suggested that women who start HT near menopause had a decreased risk of coronary heart disease when taking estrogen alone (relative risk [RR] = 0.66; 95% CI, 0.54-0.80) or in combination with progestin (RR = 0.72; 95% CI, 0.56-0.92). A secondary analysis conducted by the investigators involved in the WHI revealed that risk for coronary heart disease was not significantly increased in women under age 60 years of age or within 10 years of menopause. Further studies to evaluate the safety and efficacy of HT during the MT and the early postmenopausal years are ongoing.

Concerns about the risks of HT have increased interest in nonhormonal alternatives for the treatment of symptoms in the MT. In some women, vasomotor symptoms during the MT can be reduced by wearing layered clothing, avoiding caffeine and alcohol, and by keeping the ambient temperature a few degrees cooler. Herbal treatments, such as black cohosh (Remifemin Menopause; Enzymatic Therapy, Green Bay, WI), have been shown to have marginal or no benefit in placebo-controlled trials. Neuroactive agents, including selective serotonin reuptake inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs), alpha adrenergic agents, and others all have some efficacy in the treatment of vasomotor symptoms (Table 13–8). Both SSRIs and SNRIs may be effective because norepinephrine and serotonin appear to be involved in the hypothalamic regulation of temperature homeostasis and to play a role in the development of

TABLE 13–8 Treatment for hot flashes.

Drug	Name	Dose (mg/day)
SSRI	Citalopram	10-20
	Fluoxetine	10-20
	Sertraline	25-50
	Paroxetine	10-20
SNRI	Venlafaxine	37.5-150
	Desvenlafaxine	50-100
Alpha adrenergic agents	Clonidine	0.1
Gamma aminobutyric acid	Gabapentin	300-900

hot flashes. Randomized placebo-controlled trials have shown that SSRIs (citalopram, sertraline, paroxetine) and SNRIs (venlafaxine) can help to reduce the severity and frequency of hot flushes. Clonidine (an alpha adrenergic agonist) and gabapentin also have some efficacy. Gabapentin is the only agent compared head-to-head with estrogen suggesting equivalent efficacy when used in high doses.

The WHI study did not address the effect of hormone treatment on hot flushes and vaginal atrophy. Clearly, there are alternatives for the treatment of osteoporosis and cardiovascular disease that are superior if prevention of both conditions is the sole reason for HT. Every woman should discuss with her caregiver the optimal management for her as an individual. This should take into account medical and family history as well as symptomatology. It can be uniformly recommended, however, that menopausal women maintain appropriate nutrition, weight reduction, and exercise along with moderation in alcohol and caffeine intake and cessation of smoking.

INFERTILITY

Infertility is defined as the inability of a couple to conceive after 1 year of frequent unprotected intercourse without contraception. This definition is based on observational data showing that approximately 85% of couples achieve pregnancy after 1 year of unprotected coitus. Using this definition, approximately 15% of reproductive-age couples experience infertility. However, the diagnosis of infertility does not mean that they cannot conceive—a more precise diagnosis term would be **subfertility**, or a diminished capacity to conceive. The actual probability of the fertility potential of a population may be better assessed with variables that can quantify a monthly cycle rate. The concepts that have been used for quantitative analysis are **fecundability** and **fecundity**. Fecundability is defined as the probability of achieving a pregnancy within one menstrual cycle, and in normal couples the chance of conception after 1 month is approximately 25%. Fecundity is a related concept that is defined as the ability to achieve a live birth within one menstrual cycle.

In the United States, demands for infertility treatment have dramatically increased. The 1995 National Survey of Family Growth reported that 9.3 million women received infertility treatment in their lifetime compared with 6.8 million in 1988. This rise in treatment is due not only to increased public awareness—it also reflects the significant demographic, societal, and economic changes in our society. These include the aging of the baby boom generation, which has increased the size of the reproductive-age population. Perhaps more important is the increased use of contraception and postponement of childbearing until the last two decades of a woman's reproductive life. Approximately 20% of women in the United States now have their first child after 35 years of age.

Age alone has a significant impact on fertility and affects a woman many years before the onset of menopause. One factor is the age-dependent loss of ovarian follicles (see later). A 38-year-old woman has 25% of the fecundability of a woman under 30 years of age. Another age-related subfertility factor is that the spontaneous abortion rate increases with advancing age. The overall incidence of clinical abortion increases from 10% in women under age 30 to more than 40% in women over 40. The increased pregnancy loss can be largely attributed to abnormalities in the aging oocyte; older follicles have an increased rate of meiotic dysfunction, resulting in higher rates of chromosomal abnormalities.

The main causes of female subfertility can be classified in the following way: (1) ovulatory defects, (2) pelvic disorders, and (3) male factors. These factors account for 80% to 85% of couples diagnosed with infertility. They are not mutually exclusive—about 15% of couples have more than one cause of subfertility. In approximately 20% of couples, the cause remains unknown and is classified as unexplained infertility. This section will briefly discuss the causes of subfertility and review the diagnosis and management.

DIAGNOSIS OF INFERTILITY

Ovulatory Defects

Ovulatory disorders are responsible for 25% of cases of infertility. Ovulatory status can be obtained from the history. If a woman experiences cyclic, predictable menses at monthly intervals, ovulation can be predicted 98% of the time. This is not an invariable rule, however, and irregular menstrual cycles are not a sure sign of anovulation.

The only way to confirm ovulation is by achieving pregnancy. However, a variety of methods can indicate that ovulation has occurred. For example, a thermal shift occurs around the time of ovulation. Prior to ovulation, morning basal body temperature (BBT) is below 98°F (36.7°C), and after ovulation the temperature increases at least 0.4°F (0.2°C) for at least 1 to 3 days. This rise in temperature reflects the progesterone that is secreted from the corpus luteum, which consequently raises the hypothalamic set-point for BBT. The temperature rise occurs approximately 2 days after ovulation because of the time and dose required for the progesterone effect at the hypothalamus. BBT can, therefore, not be utilized to prospectively predict ovulation. Measurements of midluteal serum progesterone concentration can also be performed to document the occurrence of ovulation. If this level is

greater than 3 μg/L (9.5 nmol/L), it is a strong indication that ovulation has occurred. An endometrial biopsy can confirm ovulation. It is performed during the luteal phase and gives a qualitative assessment of ovulation, because the duration of progesterone exposure produces predictable endometrial histology. Lastly, a sonographic examination documenting a decrease in follicle size—or disappearance altogether of the previously developed follicle—is suggestive of ovulation. All of these methods indicate that ovulation has occurred.

There are only a few ways to predict that ovulation is going to occur. The most common way is to detect the LH surge. Ovulation typically occurs 34 to 36 hours after the onset of the LH surge.

Although ovulation may occur, some women may have a luteal-phase defect. This is characterized by an inadequate quantity or duration of progesterone secretion by the corpus luteum. There is a distinct window of time for implantation. The theory is that the progesterone deficiency desynchronizes ovulation (egg) and implantation (endometrium). However, the incidence of luteal phase defects is difficult to assess because the definition is not standardized. Typically, the diagnosis is established by luteal phase endometrial dating; if the histologic development of the endometrium lags more than 2 days beyond the day of the cycle, it is diagnostic of a luteal phase defect. However, up to 30% of women with normal cycles meet this criterion. Another method involves measuring midluteal progesterone levels. If progesterone is less than 10 ng/mL, it suggests a luteal phase defect. This is not reliable because progesterone is intermittently secreted, and the serum progesterone level can change from 1 hour to the next in the same individual. Furthermore, the lack of properly validated tests places the existence of luteal phase defects and their association with subfertility in question. Lastly, as empiric treatment for unexplained infertility has developed, delaying treatment (see later, superovulation with IUI) for exact diagnoses has less importance. More sophisticated testing of endometrial proteins required for implantation (eg, integrins, glycodelin) may revive enthusiasm for making a specific diagnosis in the future.

The cause of ovulatory dysfunction has been previously discussed. All anovulatory patients should have determination of PRL and TSH levels—and, if necessary, androgen levels—to identify the cause of the ovulatory disturbance. Treatment should be directed toward the cause, which is discussed in greater detail later. It could be argued that all patients should have evaluation of early follicular cycle (days 2-4) FSH, LH, and estradiol to assess ovarian reserve (age).

Pelvic Disorders

Pelvic disorders account for more than 30% of couples with the diagnosis of infertility. Uterine tube damage and adhesion formation are responsible for most pelvic pathologic processes causing infertility, whereas endometriosis is the primary pelvic disorder causing subfertility. The causes of tubal damage and adhesions include postinfectious state (pelvic inflammatory disease [PID]), endometriosis, and a history of pelvic surgery (especially surgery for ruptured appendicitis).

PID is defined as infection of upper genital tract structures and is usually caused by a sexually transmitted disease. The known initiating organisms are chlamydiae and *Neisseria gonorrhoeae*. The symptoms are variable but usually include lower abdominal pain, nausea, and vaginal discharge. However, in up to 30% of chlamydial infections, PID may be clinically inapparent and may remain undiagnosed until presenting with subfertility.

Endometriosis is the presence of endometrial glands and stroma outside of the uterus. There may be no manifestations other than infertility, or symptoms may progress to include severe pelvic pain, dysmenorrhea, and dyspareunia. The diagnosis is suspected if findings on surgical exploration show characteristic lesions. Lesions can be staged according to published criteria. Diagnosis is confirmed with biopsy of the peritoneal lesions. The disease occurs in approximately 3% to 10% of reproductive-age women and may be responsible for up to 25% to 35% of the female factors responsible for subfertility.

The pathogenesis of endometriosis is not completely known. A prominent theory (Sampson) involves retrograde menstruation. It is well established that menses can flow through the uterine tubes into the abdominal cavity. In fact, this phenomenon is thought to occur in almost all menstruating women. There is good evidence that the endometrial tissue subsequently invades and proliferates into the peritoneum. It is theorized that the immune system should normally dispose of the tissue and that altered immunity may result in implantation of this endometrial tissue outside the uterus in those women who subsequently develop endometriosis.

There is a strong association between adhesive disease and endometriosis. In these cases, the cause of subfertility is a result of distorted anatomy and consequently altered function. However, in mild cases of endometriosis, where only peritoneal lesions are identified and no anatomic distortion exists, the cause of infertility remains uncertain and controversial. There is evidence that the peritoneal fluid is altered in the presence of endometrial tissue with increased macrophages and inflammatory mediators. Several studies suggest that inflammatory changes result in adverse effects on folliculogenesis, ovum transport, fertilization, and implantation.

Tubal damage can be diagnosed with a hysterosalpingogram or surgical exploration. Hysterosalpingography involves the introduction of radiopaque contrast media into the pelvis through the cervix and then fluoroscopy, revealing an outline of contrast in the uterine cavity, uterine tubes, and peritoneal cavity. The diagnosis of pelvic endometriosis can only be made by surgery, and most often the surgical procedure is laparoscopy. Diagnostic laparoscopy is usually done when there is a high suspicion of endometriosis based on the clinical history or adhesive disease based on a history of PID or pelvic surgery. Laparoscopy may also be performed, if all other tests are normal and the couple continues not to achieve a pregnancy.

Male Factor Causes

The male factor contributes in 40% to 50% of cases of diagnosed infertility, and all evaluations should include the male partner. The diagnostic test for male factor infertility is the semen analysis.

Although this is a largely descriptive test (volume, sperm count, motility, and morphology), there is some correlation with pregnancy outcome. This should be used solely as a screening test. To understand the pathophysiology of male factors, it is important to review the physiology of spermatogenesis and the anatomy of the male reproductive tract as described in Chapter 12.

Unexplained Infertility

Approximately 15% to 20% of the couples diagnosed with infertility have no identifiable cause after a full investigation. The workup should include documentation of ovulation (regular menstrual cycles, progesterone levels, or LH testing), hysterosalpingography to verify patent fallopian tubes, and a semen analysis. The term *unexplained* implies that there is a potential explanation for the subfertility, but the cause has yet to be identified. The cause may be subtle abnormalities in folliculogenesis, sperm-ovum interactions, or defective implantation.

Several studies have evaluated the natural history of unexplained infertility. It is estimated that fecundity in younger couples (female partner under the age of 40) with unexplained infertility is 3% to 5% compared with 20% to 25% in the age-matched couples with normal fertility. Treatment involves methods that increase fecundability and are discussed later.

MANAGEMENT OF THE INFERTILE COUPLE

It is important to remember that in most couples, there is a chance for spontaneous conception. Recent studies estimate that the average probability for live birth without treatment is 25% to 40% during the 3 years after the first infertility consultation. This translates into a cycle fecundity rate of 0.7% to 1% per month. The presence of endometriosis, abnormal sperm, or tubal disease independently reduces the chance of spontaneous pregnancy and live birth by approximately 50% for each variable. Infertility for more than 3 years, female age over 30 years, and primary infertility were important negative prognostic factors.

Evaluation should focus on known causes of infertility or subfertility: ovulatory defects, pelvic disorders (tubal disease, endometriosis), and male factor issues.

Ovulatory Disorders

Treatment should be diagnosis-specific, if possible. For the female, this means that the cause of any ovulatory defect should be determined and specific treatment then instituted. This enhances outcome and decreases the risk of complications (spontaneous abortion and multiple gestation). This treatment may include the use of dopamine agonists (hyperprolactinemia), thyroid replacement (hypothyroidism), pulsatile GnRH (hypogonadotropic hypogonadism), or clomiphene citrate (for PCOS). The most common cause of anovulation is inappropriate feedback such as in PCOS. Ovulation in patients with PCOS can be induced with clomiphene citrate, a nonsteroidal agonist-antagonist of estrogen that blocks the hypothalamic-pituitary axis from feedback by circulating estrogens. As a result, there is increased gonadotropin release to stimulate follicular recruitment and ovulation. In addition, because PCOS is associated with insulin resistance, and elevated insulin directly has an impact on the ovary, insulin sensitizers such as metformin have been used to enhance ovulatory response in women with PCOS. However, a large multicenter randomized trial compared clomiphene to metformin and found clomiphene to be significantly more effective for ovulation induction and achieving a live birth Most recently, letrozole, when compared with clomiphene, was associated with higher live-birth and ovulation rates among infertile women with PCOS.

Pelvic Disorders

In general, adhesive tubal lesions should be treated surgically. The location and extent of disease should, however, be evaluated. Patients with distal tubal occlusion—unless it is very mild—are most often better served by assisted reproduction (*in vitro* fertilization [IVF]). Other possible causes of infertility should also be examined. If a patient has a history of documented tubal disease, in addition to other abnormalities (ovulatory dysfunction or male factor infertility), or if they are over 35, the likelihood for successful surgical management decreases by approximately 50%, and consideration for avoiding surgery and moving directly to assisted reproduction is paramount. The exception to this rule is the presence of hydrosalpinges on ultrasound. The presence of hydrosalpinges that retain fluid when nondistended (ie, not seen with hysterosalpingography) leads to a significant reduction in outcome with assisted reproductive therapy (ART). Prior removal or proximal occlusion of the tube to prevent contamination of the uterine cavity should be performed before ART is offered.

There are conflicting data in the literature concerning the appropriate treatment for mild endometriosis. A well-designed randomized trial from Canada evaluated the effect of surgical treatment on pregnancy outcome for patients diagnosed with mild endometriosis without anatomic distortion. It showed that pregnancy rates at 9 months post-laparoscopy were 27% in the surgically treated group compared with 18% in the untreated group. Severe endometriosis (disease that alters the pelvic anatomy or involves the ovary with endometriomas) should be surgically treated to restore normal pelvic anatomy. There appears to be no advantage to medical therapy for endometriosis in women seeking fertility.

Male Factor Infertility

Male factor infertility is discussed in Chapter 12. Like female partner treatment, therapy should be targeted, if possible, toward the cause of subfertility. Obstructive disease may be treated surgically. A prominent varicocele with a stress pattern on semen analysis (decreased motility with an increase in abnormal morphology) may suggest a need for surgical repair. Any endocrinologic abnormalities (while less common in the male) should be treated (eg, prolactinoma). Unfortunately, beyond this point, most treatments require a combined approach very similar to that discussed later for unexplained infertility.

Unexplained Infertility

The treatment of unexplained infertility can be frustrating for the physician and the patients because the recommendations for therapy are not targeted toward a specific diagnosis. Although there are very limited evidence-based data to guide treatment, therapy should be directed toward increasing the fecundability rate. The two main treatments are superovulation plus intrauterine insemination (IUI) and IVF.

Superovulation methods are designed to qualitatively improve the cycle and to hyperstimulate the ovary with rescue of more follicles (quantitative improvement). Administration of clomiphene citrate or of gonadotropins can be used for superovulation. However, success rates are only significantly higher than expectant management when combined with IUI of washed sperm. IVF provides a higher fecundability rate. However, these procedures are significantly more costly and more invasive, and they should be used after a trial (3 or 4 cycles) of superovulation and IUI has failed. Evidence suggests that persistent treatment with IUI is more expensive and time to live birth is increased, compared to treatment with IVF. With older patients, aggressive therapy should be considered earlier in the treatment effort.

CONTRACEPTION

Approximately 50% of all pregnancies in the United States are unplanned. In adolescents and in women of older reproductive age, the unplanned pregnancy rate is higher, approaching 82% and 77%, respectively. This equates to 3.5 million unplanned pregnancies in the United States per year. The advent of oral contraceptives has meant that women have been able to postpone childbearing. However, approximately 50% of unplanned pregnancies are due to contraceptive failures. Possible causes of failure include lack of education, poor compliance, and side-effect profiles. In this section, we discuss the different methods of hormonal contraception. We shall discuss also the target population for the various modalities of contraception and their mechanisms, side effects, and ways to decrease failure rates. At the end of the section, emergency contraception will be discussed briefly.

ORAL CONTRACEPTIVES

Combination Pills

In the United States, oral contraceptive pills are the most widely used method for contraception. There are two types of oral contraception: combination pills and progestin-only pills. The various hormones used in birth control pills are illustrated in Figure 13–19.

The development of oral contraceptive agents began with the isolation of progesterone. However, progesterone was very expensive and difficult to isolate. Ethisterone, a derivative of an androgen, was found to have progestin activity and was much easier to isolate than progesterone. With removal of carbon 19, the progestational activity was increased, and the new compound was termed norethindrone. When this hormone was administered to

women, ovulation was inhibited. During the process of norethindrone purification, an estrogen contaminant was found. When this contaminant was removed, women would experience breakthrough bleeding. The estrogen was added back, thereby creating the first-generation combination birth control pill, which was approved by the Food and Drug Administration (FDA) in 1960.

Oral contraceptives can be divided into generations based on dose and type of hormone. The first-generation birth control pills contained more than 50 μg of ethinyl estradiol or mestranol and a progestin. The adverse events associated with high-dose estrogen, such as coronary thrombosis, led to development of the second-generation pill, which contained less than 50 μg of ethinyl estradiol and progestins other than levonorgestrel derivatives. Next, attention was directed toward the progestin, which was thought to have adverse androgenic effects such as affecting lipid profiles and glucose tolerance. This led to the development of third-generation pills that contained both a lower dose of estrogen (20-30 μg of ethinyl estradiol) and newer progestins (gonanes: desogestrel or norgestimate). Indeed, studies have demonstrated a reduction in metabolic changes associated with these progestins, but limited data are available to show any actual reduction of cardiovascular events. Another recently developed progestin, drospirenone, has antimineralocorticoid and antiandrogenic activity in addition to its pharmacologic progestational effects. As an analog of spironolactone rather than androgen, it competitively binds to aldosterone receptors, and it may counteract the estrogen stimulation of the renin-angiotensin system, resulting in more weight stability and less water retention. A newer combination oral contraceptive, Yasmin, which contains 3 mg drospirenone and 30 μg ethinyl estradiol, has been approved by the FDA and is prescribed for women with hyperandrogenism or other side effects attributable to oral contraceptives. However, the relative antiandrogenic activity of drospirenone is small compared with cyproterone acetate or the therapeutic dose of spironolactone used for the treatment of hirsutism.

Contraceptives can be classified also based on formulas or schedules of administration. The theory behind phasic preparations was to further decrease the amount of total progestin administered in an attempt to reduce metabolic changes attributed to the progestin, thereby decreasing adverse effects. The traditional monophasic pill (eg, Loestrin) contains 30 μg of ethinyl estradiol and 1.5 mg of norethindrone. This dose is given every day for 3 weeks with a 1-week, hormone-free interval. The progestin dose remains constant throughout the cycle. The second type is the biphasic pill (eg, Ortho-Novum 10/11), which contains 35 μg of ethinyl estradiol and either 0.5 or 1 mg of norethindrone. The 0.5 mg of norethindrone is administered in the first 10 days of the month, and the 1 mg is administered for the following 11 days. The last 7 days of the cycle are free of hormone. With this combination, there was a theoretical increase in breakthrough bleeding and an increased pregnancy rate. A meta-analysis revealed no difference, but limited data were available.

Because of concerns that this regimen might result in both breakthrough bleeding and pregnancies, another phasic formulation was developed. The triphasic pills (eg, Triphasil, Ortho-Novum 7/7/7) contain 0.5, 0.75, and 1 mg norethindrone

combined with 35 μg ethinyl estradiol. Theoretically, this formulation improves cycle control. There are several other regimens, some of which alter estrogen doses to simulate the estrogen cyclic rhythm (Triphasil-30, Triphasil-40, Triphasil-30 μg ethinyl estradiol) and possibly decrease breakthrough bleeding. A meta-analysis comparing biphasic versus triphasic pills revealed that triphasic pills significantly improved cycle control. However, the progestins in each pill tested were different, and this could account for better cycle control rather than the phasic formulation. An additional meta-analysis was performed on triphasic versus monophasic pills to assess cycle control and metabolic effects. This analysis revealed no difference between the formulations. Therefore, there is little

scientific rationale for prescribing phasic preparations in preference to the monophasic pill.

The pharmacologic activity of progestins is based on the progestational activity and bioavailability of each progestin as well as the dose. The relative potencies of the different progestins are levonorgestrel greater than norgestrel greater than norethindrone. The active estrogen component of oral contraceptives is ethinyl estradiol (even if mestranol is administered).

When the hormones are administered for 21 days of the cycle, there is enough progestin to inhibit rapid follicle growth for about 7 more days. Figure 13–20 demonstrates that during the steroid-free interval there is no rise in estrogen, indicating no follicular
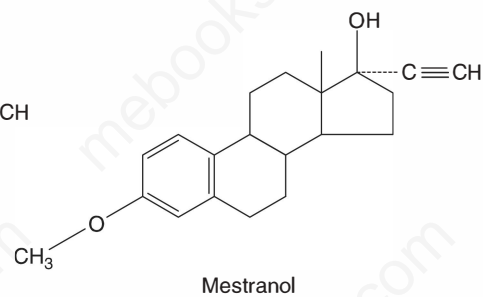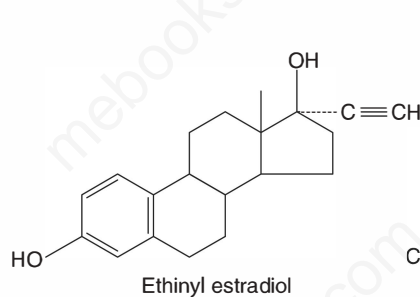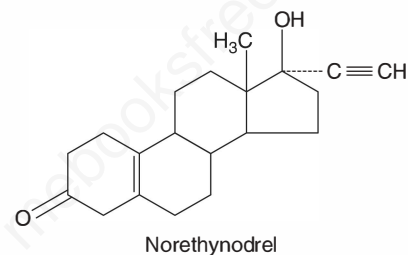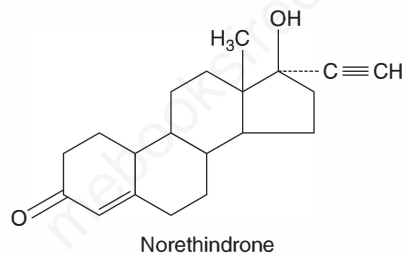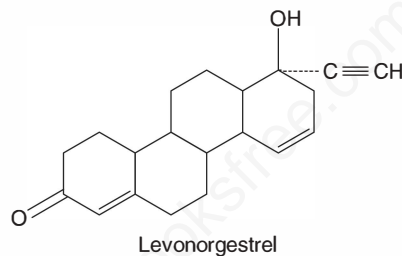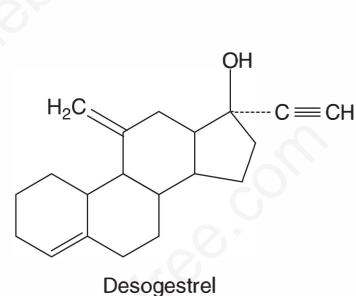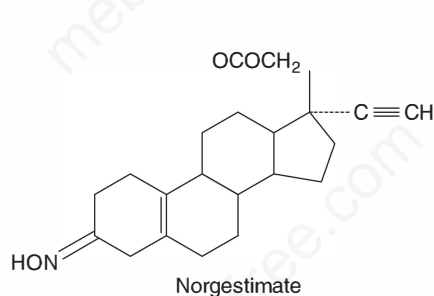
FIGURE 13–19 Oral contraceptive pill hormonal components. All except endogenous hormones are synthetic steroids.

Pregnanes

Medroxyprogesterone acetate
(Provera)

Endogenous hormones

Estradiol

Progesterone
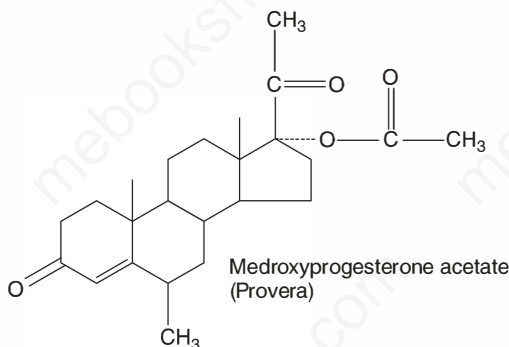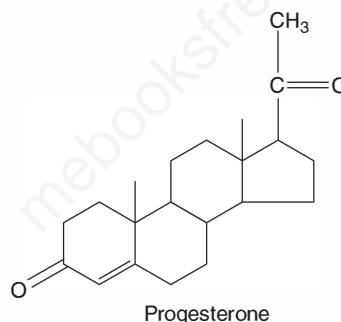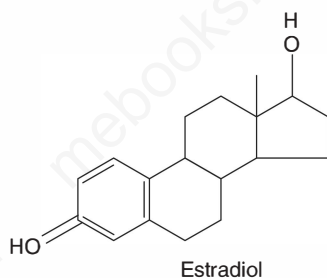
FIGURE 13–19 (*Continued*)

maturation. It is likely that pills missed after this time are responsible for some of the unintended pregnancies. Therefore, it is important that this interval not be extended.

Pharmacologic doses of progestin inhibit ovulation by suppressing GnRH pulsatility and possibly inhibiting release of pituitary LH. Progestins also impair implantation and produce thick, scanty cervical mucus that retards sperm penetration. These latter methods play a minor role in the mechanism of oral contraception.

Ethinyl estradiol helps prevent the selection of a dominant follicle by suppressing pituitary FSH. In addition to FSH suppression, ethinyl estradiol provides stability to the endometrium, decreasing breakthrough bleeding. It also upregulates the progesterone receptor and decreases clearance, thereby potentiating the activity of the progestin.

Traditionally, a pill is administered daily for 3 weeks out of 4, preferably at the same time each day (more critical with progestin-only pills). This regimen was designed to mimic the menstrual cycle with monthly withdrawal bleeding. The conventional start date is on the first Sunday after menses (first day of menses for triphasics). An alternative method is to start at the time of the clinic visit, regardless of the day of the menstrual cycle, with a backup method for 7 days (*Quick-Start*). This method has the advantage of immediate contraception without adverse bleeding events. In addition, with the Quick-Start method, women are more likely to start the second pack of oral contraceptives, suggesting increased compliance. During the 28-day regimen, there is a 1-week, steroid-free interval. If the steroid-free interval is prolonged beyond the 7-day window, ovulation is possible. Therefore,

this is a critical time not to neglect taking pills. On the other hand, a woman may continue to take the hormone pills and skip the steroid-free interval to avoid monthly bleeding. A randomized trial comparing continuous oral contraception with the traditional cyclic method revealed a significantly greater incidence of erratic bleeding with (overall) the same number of days of bleeding, albeit with a reduced volume of bleeding. New formulations have been developed for those who do not desire cyclic bleeding. This regimen involves 84 days of continuous hormone administration followed by a steroid-free interval of 1 week. This can easily be done with the traditionally packaged oral contraceptive pills.

A routine for daily administration improves compliance and contraceptive efficacy. Failure to take the pill at the same time every day and not understanding the package insert are associated with missing two or more pills during the cycle. To decrease failure rates, women should understand that if they forget to take the pill, they must use barrier prophylaxis.

Postpartum women who are not breastfeeding may begin combination oral contraceptives 3 weeks after delivery. For women who are breastfeeding, it is advised that institution of combination oral contraceptives be delayed until 3 months postpartum. The recommendation for this delay is due to decreased milk letdown secondary to estrogen but may be waived once lactation is well established.

Noncompliance increases the incidence of unwanted pregnancies. Appropriate use of birth control is achieved 32% to 85% of the time in the general population. Teenagers have at most a 50% continuation rate, and 25% of pill users discontinue the practice in the first year. The efficacy of the oral contraceptive under
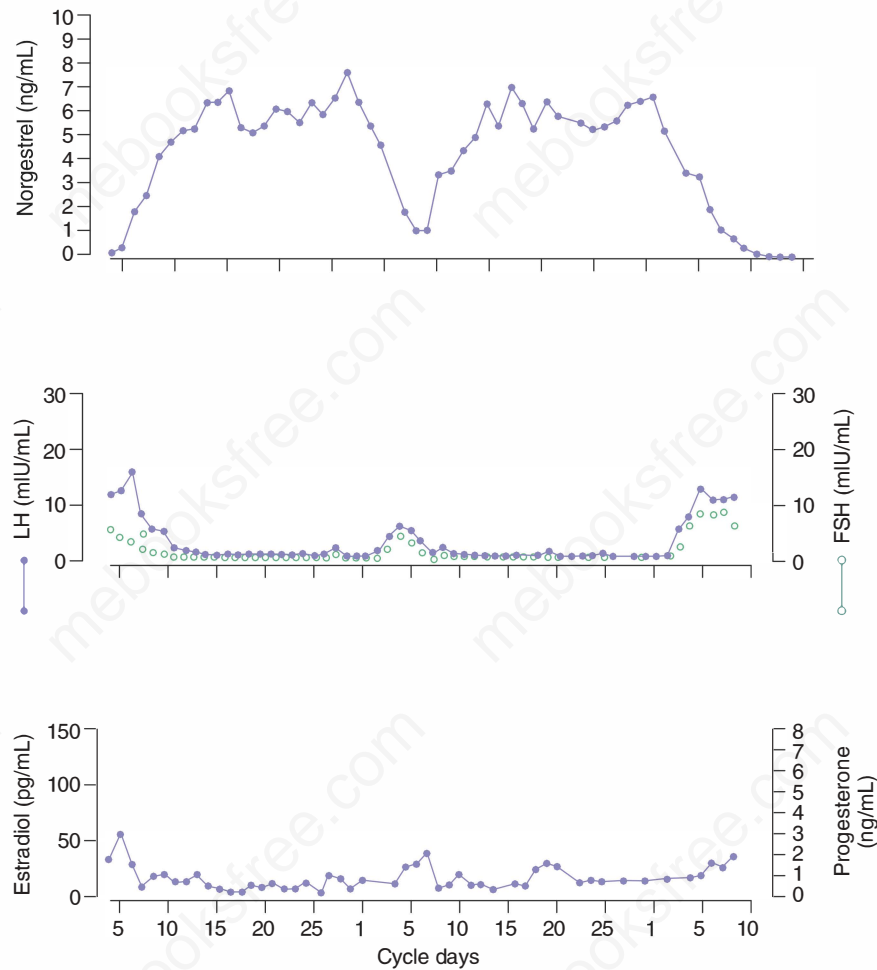
FIGURE 13–20 Progestin effects on steroidogenesis and ovulation. (Reproduced with permission from Brenner PF, Mishell DR Jr, Stanczyk FZ, et al. Serum levels of d-norgestrel, luteinizing hormone, follicle-stimulating hormone, estradiol, and progesterone in women during and following ingestion of combination oral contraceptives containing dl-norgestrel. *Am J Obstet Gynecol*. 1977 Sep 15;129(2):133-140.)

conditions of perfect use is 0.1 failures per 100 woman-years (0.1 per 100 users). With typical use, the failure rate is 3%, with first-year failure rates approaching 7.3% to 8.5%. Side effects contribute to noncompliance. The most common side effect is breakthrough bleeding. Other unwanted symptoms include bloating, breast tenderness, nausea, and possibly headaches, weight gain, and depression. Some studies suggest that altering estrogen doses may improve symptoms. Failure rates may also be associated with concomitant use of drugs (eg, rifampin, hydantoins) that accelerate hormone metabolism.

A. Other benefits There are noncontraceptive benefits to the pill. These include reduced monthly blood loss (less iron deficiency) and less dysmenorrhea as well as reduced benign breast disease and mastalgia. Oral contraceptives also reduce the incidence of PID and ectopic pregnancies. Other significant benefits include a reduction in risk of ovarian cancer, endometrial cancer, and colorectal cancer. Oral contraceptive use also has cosmetic benefits; it can improve excess hair growth and acne (see PCOS). It is not unusual for the pill to be administered for noncontraceptive problems.

B. Potential risks In general, oral contraceptives have proved to be safe for most women, but the possibility of adverse effects has received much attention. Unfortunately, the literature is full of conflicting reports. Data concerning controversial adverse effects will be discussed in the following section.

Although the estrogen in combination oral contraceptives tends to increase triglycerides and total cholesterol, these levels are still within the normal range, and the contraceptives appear to increase HDL and decrease LDL. Progestins attenuate these effects, which suggest an adverse metabolic milieu. Although it is known that low HDL-LDL ratios are associated with cardiovascular events, the importance of lipid changes associated with birth control pills is unknown. To date there is no strong evidence of an increased incidence of myocardial infarctions in healthy nonsmoking oral contraceptive users. In patients with other cardiovascular risk factors such as hypertension—at least in Europe—there is an up to 12-fold increased risk of cardiovascular events. Women who are over 35 years of age and who smoke are at increased risk of cardiovascular events. This risk is amplified with use of the birth control pills. If more than 15 cigarettes per day are smoked, there

is a relative risk (RR) of 3.3 for a cardiovascular event, compared with an RR of 20.8 with concomitant use of oral contraceptives. However, if a woman smokes fewer than 15 cigarettes per day, there is an RR of 2.0 for a cardiovascular event, compared with an RR of 3.5 with concomitant contraceptive use. Former smokers after 1 year have no significant increased risk. First-generation oral contraceptives (>50 μg ethinyl estradiol) imposed an RR for stroke (ischemic or hemorrhagic) of 5.8. With low-dose agents (<50 μg ethinyl estradiol), there appears to be no significant increased risk of stroke among healthy normotensive nonsmoking women. The RR associated with hemorrhagic stroke for hypertensive women is 10.2 to 14.2. Because the potential exists for adverse outcomes, women taking oral contraceptives should be screened regularly for cardiovascular risk factors to ensure safe administration.

The risk of deep vein thrombosis and pulmonary embolism is increased two- to threefold with administration of the pill. The mechanism by which oral contraceptives enhance venous thrombosis is unknown, but there may be estrogen-related changes in coagulation parameters. These include increased clotting factors and activation of platelets and a decrease in protein S and fibrinolytic activity. However, these changes in measured serum clotting factors do not predict the occurrence of deep vein thrombosis. Genetic thrombophilias increase the risk of venous thrombosis. The prevalence of factor V Leiden in the general population is 5%. The incidence of deep vein thrombosis among this population is 60 per 100,000 per year, and with use of oral contraceptives, the incidence approaches 280 to 300 per 100,000 per year. The baseline incidence of deep vein thrombosis in women is approximately 3 per 100,000 per year, and with current oral contraceptive uses it is 9.6 to 21.1 per 100,000 per year. For comparison, during pregnancy, the incidence of deep vein thrombosis is 60 per 100,000 per year. Older age (>40-44) increases the incidence two- to threefold but does not affect the RR. There is no evidence that smoking has an effect on the incidence of deep vein thrombosis with oral contraceptive use. At this time, universal screening for thrombophilias is not cost effective. However, any history of deep vein thrombosis warrants a workup for thrombophilia.

The association of oral contraceptives and cancer risk has been evaluated for breast, cervical, and liver cancer. Observational studies investigating a possible association between oral contraceptive use and breast cancer have reported conflicting results. The most recent information is that oral contraceptive usage (current or past use) has no impact on the incidence of breast cancer—RR 1.0 (CI 95% 0.8-1.3)—among women 35 to 64 years of age. Several observational studies have linked oral contraceptive use with invasive cervical cancer, although it is not clear if this association is causally related. A recent study investigating the association between oral contraception use and cervical cancer revealed a nearly threefold increased risk among human papillomavirus carriers with 5 to 9 years of oral contraceptive use (RR 2.82; 95% CI 1.46-5.42). This evidence suggests that women taking oral contraceptives should be screened yearly with Pap smears to prevent cervical cancer. In the 1980s, there was an association between hepatocellular carcinoma in women under 50 years of age and oral contraceptive use. With further investigation there appears to be

TABLE 13–9 Contraindications to combination oral contraceptive use.

Pregnancy
Increased thromboembolic risk
Smoking and age ≥35 years
Hepatoma
Breast cancer
Endometrial cancer
History of cerebrovascular or cardiovascular disease
Undiagnosed vaginal bleeding
Uncontrolled hypertension

no increased risk of hepatic cancer with the use of oral contraceptives.

After discontinuation of oral contraceptives, the activity of the HPO axis gradually returns to a precontraceptive state. After a 2- to 4-week prolongation of the follicular phase, the LH peak is observed, which suggests that the suppressive effects of the oral contraceptive have dissipated and that cyclic menses will resume.

Contraindications to oral contraceptive administration are summarized in Table 13–9.

Progestin Only

Progestin-only pills (Ortho Micronor, Nor-QD, 0.35 mg norethindrone; Ovrette, 0.075 mg levonorgestrel) are also available for contraception. The target population for administration of progestin-only contraception includes women with contraindications to estrogen, breastfeeding mothers, and older women.

The circulating levels of progestin following ingestion of progestin-only pills (minipill) are 25% to 50% of that following ingestion of estrogen-progestin oral contraceptive pills. Serum levels that peak 2 hours after administration are followed by rapid elimination (Figure 13–21). The peak levels of norethindrone and levonorgestrel vary (4-14 and 0.9-2 ng/mL, respectively), but 24 hours after pill ingestion, serum levels are 0.2 to 1.6 ng/mL and 0.2 to 0.5 ng/mL, respectively. Thus, there is no accumulation of progestin over time. Progestin-only administration results in lower steady state levels and a shorter half-life compared with concomitant administration with estrogen.

Owing to the lower levels of progestin in the progestin-only pill, there is less influence on the inhibition of ovulation and more impact of thickening cervical mucus (*hostile environment*) inhibiting sperm penetration. Sperm that are able to penetrate have decreased mobility. Progestins also alter the endometrial lining (inhibition of progesterone receptor synthesis and reduction in endometrial glandular development, preventing implantation) and perhaps inhibit the motility (number and motility of the cilia) of the uterine tube. LH peaks—as well as FSH peaks—are suppressed compared with pretreatment levels. The change in cervical mucus takes place 2 to 4 hours after the first dose. However, after 24 hours, thinning of the cervical mucus is evident, allowing unimpaired sperm penetration, which is why it is critical to take the progestin-only pills at the same time every day.

The progestin-only pill should be started on the first day of menses. This pill should be taken at the same time every day. If

A. Norethindrone

B. Levonorgestrel
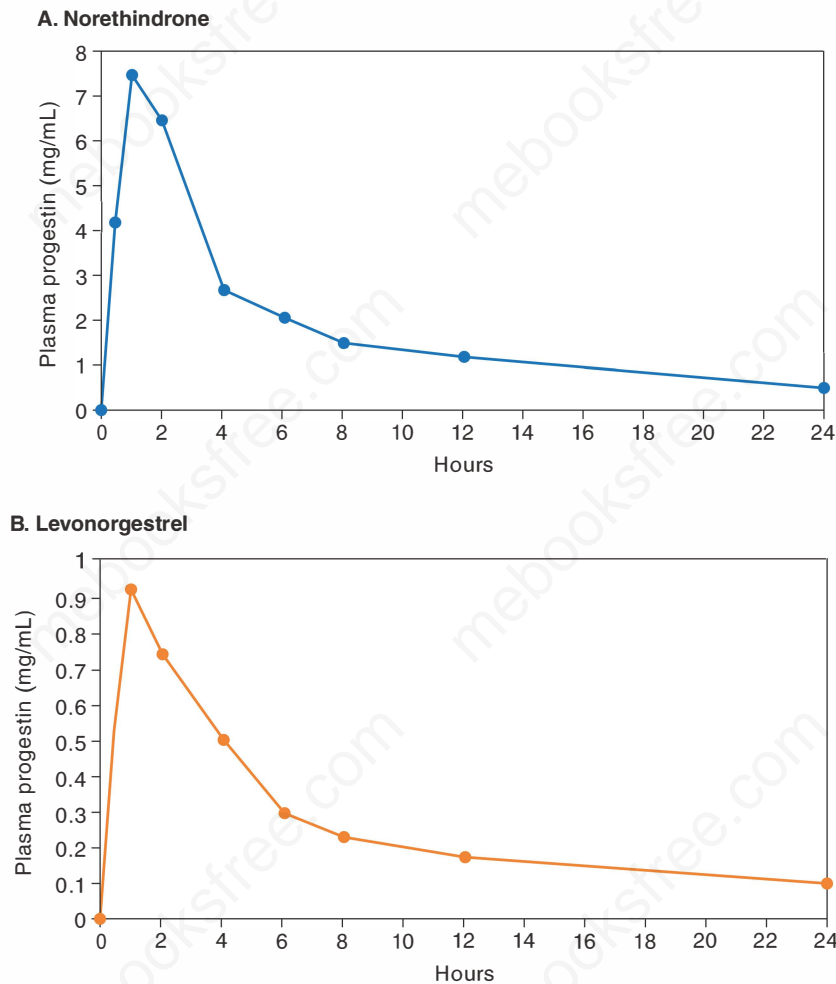
FIGURE 13–21 Serum levels attained with progestin-only pills. [Reproduced with permission from McCann MF, Potter LS. Progestin-only oral contraception: a comprehensive review. *Contraception.* 1994 Dec;50(6 Suppl 1):S1-S195.]

administration is 3 hours late, a backup method should be used for 48 hours. If a pill is missed, a backup method should be used for 48 hours, and if two or more pills are missed, a backup method should be used for 48 hours due to rapid resumption of the cervical mucus effect. If no menses occur in 4 weeks, a pregnancy test is necessary. Progestin-only pills may be administered immediately postpartum.

The efficacy of progestin-only pills under conditions of perfect use is 0.3 to 3.1 failures per 100 woman-years (failure rates of 1.1%-9.6% in the first year) or 0.5 per 100 users. This efficacy rate is achieved only with careful compliance. The typical use is associated with a greater than 5% failure rate. Failure rates were lowest in women over 38½ years of age and those who were breastfeeding. The efficacy may also be influenced by body weight and by concomitant use of anticonvulsants. The major disadvantage is that the pill must be administered at the same time every day. As a result of even slight flexibility in the schedule, there is increased contraceptive failure.

The risks associated with progestin-only pills are minimal. Various studies have revealed no significant impact on lipids, carbohydrate metabolism, blood pressure, or the incidence

of myocardial infarction and stroke. Furthermore, no adverse coagulation parameters have been associated with its use. There are almost no data on the association of progestin-only pills and endometrial, ovarian, cervical, or breast cancer. The major side effect is breakthrough bleeding (40%-60%). Other side effects include acne and persistent ovarian cysts. With discontinuation of the pill, menses resume with no impact on subsequent pregnancy rates or future fertility.

CONTRACEPTION: LONG-ACTING CONTRACEPTIVES

The high rate of unintended pregnancies has led to the development of long-acting reversible contraceptive modalities. Interest in long-acting methods is increasing because they offer convenience, obviate problems of compliance, and thus offer higher efficacy. Most long-acting systems contain either combination or progestin-only hormones. The effectiveness of these hormones is prolonged, mostly due to the sustained system that results in a gradual release. The modes of administration include injectables,

transdermal patches, subdermal rods, vaginal rings, and intrauterine devices (IUDs). The various types of long-acting contraceptives are discussed later.

Injectable Contraceptives

A. Progestin only   Injectable progestins that contain MPA (Depo-Provera) are beneficial when women have contraindications to estrogen, use antiepileptics, are mentally handicapped, or have poor compliance. Furthermore, there is good evidence that its use is safe in the presence of coronary artery disease, congestive heart failure, diabetes, tobacco use, and a history of venous thromboembolism. Other uses of MPA include treatment of metastatic endometrial or renal carcinoma.

Although most other long-acting contraceptives are sustained-release formulations, Depo-Provera (150 mg MPA) is provided as an aqueous microcrystalline suspension that gradually declines throughout the cycle (Figure 13–22). Pharmacologic levels (>0.5 ng/mL) are achieved within the first 24 hours and peak (at 2 ng/mL) within the first week after the injection. Serum concentrations are maintained at 1 ng/mL for approximately 3 months. Interestingly, estrogen concentration is in the early to midfollicular

level (below 100 pg/mL) and persists for 4 months after the last injection. The serum concentration of MPA decreases to 0.2 ng/mL during the last 5 to 6 months (ovulation occurs when levels are <0.1 ng/mL). However, one study observed progesterone levels to rise after 3½ months.

The mechanism of action depends mainly on the ability of higher peaks of hormone to inhibit ovulation (LH surge). Like other progestins, MPA increases cervical mucus viscosity, alters the endometrium, and decreases the motility of the uterine tubes and uterus. FSH levels are minimally suppressed with Depo-Provera.

The manufacturer's recommendation is to administer the agent every 3 months, starting within 5 days of the onset of menses and not to exceed 1 erweek. The agent is injected deeply into the upper-outer quadrant of the buttock or deltoid without massage to ensure slow release. If the subject is postpartum and not breastfeeding, Depo-Provera should be given within 3 weeks after delivery and if lactating within 6 weeks (Table 13–10).

Because compliance is not an issue, the failure rate is minimal at 0 to 0.7 per 100 woman-years (0.3 per 100 users). Weight and use of concurrent medications do not affect the efficacy. However, continuation rates are poor at 50% to 60% because of the side-effect profile. The major dissatisfaction that leads to

FIGURE 13–22 Serum medroxyprogesterone acetate (MPA) levels following injection of Depo-Provera. Open bars represent serum estradiol; closed bars represent serum progesterone concentrations. (Reproduced with permission from Ortiz A, Hirol M, Stanczyk FZ, et al. Serum medroxyprogesterone acetate [MPA] and ovarian function following intramuscular injection of depo-MPA. *J Clin Endocrinol Metab*. 1977 Jan;44(1):32-38.)

TABLE 13–10 Scheduling for injectable contraceptives.

Method	DMPA	MPA/E$_2$C
First injection		
Spontaneous menstrual cycle	Within 5 d of menses onset	Within 5 d of menses onset
Spontaneous or elective first-trimester abortion	Within 7 d	Within 7 d
Term delivery	Within 3 wk postpartum if not lactating; within 6 wk postpartum if lactating	Between 21 and 28 d postpartum if not lactating
Switching from combination OCs	With administered active pills or within 7 d after administering the pill pack's last active tablet	With administered active pills or within 7 d after administering the pill pack's last active tablet
Switching from DMPA	—	Within 13 wk after last DMPA injection
Switching from MPA/E$_2$C	Within 33 d after previous injection	—
Switching from levonorgestrel implant	Any time within 5 y of implant insertion; use of a condom back-up is recommended for 1 wk	Any time within 5 y of implant insertion; use of a condom backup is recommended for 1 wk
Switching from Copper T 380A IUD	First injection should occur before IUD removal and within 10 y after IUD insertion; a condom should be used as a back-up if the first injection is not administered within 5 d of menses onset	First injection should occur before IUD removal and within 10 y after IUD insertion condom should be used as a back-up if the first injection is not administered within 5 d of menses onset
Subsequent injections		
Injection interval	Every 12 wk or 3 mo; earlier reinjections are acceptable	Every 28 d or 4 wk or monthly; reinjection earlier than 23 d may impair cycle
Grace period	2 wk (14 wk from last injection); after 1 wk manufacturer recommends pregnancy testing before repeat injection	Control ±5 d (23-33 d from last injection); thereafter, pregnancy testing needed before repeat injection

Abbreviations: DMPA, depot medroxyprogesterone acetate; IUD, intrauterine device; MPA/E$_2$C, medroxyprogesterone acetate and estradiol cypionate; OCs, oral contraceptives.

Reproduced with permission from Kaunitz AM. Injectable long-acting contraceptives. *Clin Obstet Gynecol.* 2001 Mar;44(1):73-91.

discontinuation is breakthrough bleeding, which approaches 50% to 70% in the first year of use. Other side effects include weight gain (2.1 kg/y), dizziness, abdominal pain, anxiety, and possibly depression. Another disadvantage with the use of Depo-Provera is a delay in fertility after discontinuation. Ovulation returns when serum levels are less than 0.1 ng/mL. The time from discontinuation to ovulation is prolonged. Only 50% of patients ovulate at 6 months after discontinuing the medication, and although the agent does not cause infertility, achieving pregnancy may be delayed for more than 1 year. (The length of time for release at the injection site is unpredictable.) After the first year, 60% of women become amenorrheic, and at 5 years, the incidence of amenorrhea approaches 80%, which can be considered a potential benefit. Other benefits with use of MPA include prevention of iron deficiency anemia, ectopic pregnancy, PID, and endometrial cancer. In addition, Depo-Provera is a recommended contraceptive for women with sickle cell disease (decreased crisis) and seizure disorders (raises seizure threshold). Other therapeutic uses include dysmenorrhea and endometrial hyperplasia or cancer.

One major concern with use of Depo-Provera is the development of osteopenia, with possible advancement to osteoporosis later in life. Several observational studies have evaluated the potential impact on bone. A prospective trial revealed that current users, after 12 months of use, experience a mean BMD loss of 2.74%. However, on examining former users 30 months later, it was found that mean BMD was similar to that of nonusers, indicating that the loss is reversible and of minimal clinical importance. An ongoing multicenter study assessing BMD in users versus nonusers should clarify the impact of Depo-Provera on bone. BMD in adolescents has also been investigated because adolescence is a critical time in bone mineralization. A small prospective study revealed that BMD was decreased by 1.5% to 3.1% after 1 and 2 years of use in Norplant users, compared to increases in BMD of 9.3% and 9.5% in controls. This is a potential concern and has also led to a prospective multicenter study investigating the use of Depo-Provera in adolescents. Although one possible cause is less exposure to estrogen, an alternative and perhaps not exclusive theory involves MPA-dependent glucocorticoid activity that impairs osteoblast differentiation. Other potential risks include an adverse lipid profile (increase in LDL, decrease in HDL) and a slightly increased risk of breast cancer. The association of breast cancer with use of Depo-Provera is minimal within the first 4 years of use, with no risk after 5 years of use. Paradoxically, MPA has been used for treatment of metastatic breast cancer.

B. Combination The development of monthly combination injectables (Lunelle) has responded to the erratic bleeding associated with Depo-Provera (Figure 13–23). The cycle control is similar to what is achieved with combined oral contraceptives. The monthly withdrawal bleeding occurs 2 weeks after the injection. The target populations are adolescents and women who have difficulty with compliance. Lunelle is an aqueous solution containing
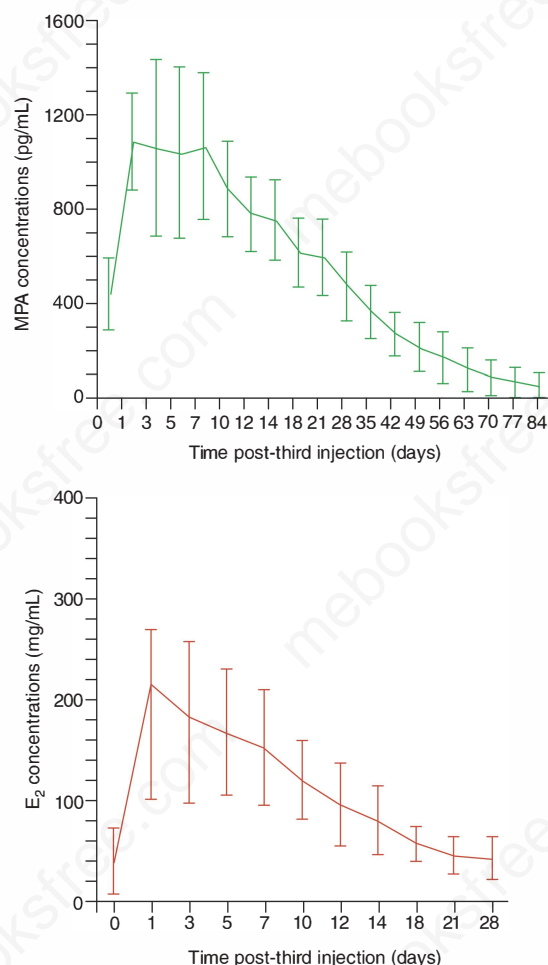
FIGURE 13–23 Serum medroxyprogesterone acetate (MPA) and estradiol levels following Lunelle injection. (Reproduced with permission from Rahimy MH, Ryan KK, Hopkins NK. Lunelle monthly contraceptive injection (medroxyprogesterone acetate and estradiol cypionate injectable suspension): steady-state pharmacokinetics of MPA and E2 in surgically sterile women. *Contraception.* 1999 Oct;60(4): 209-214.)

25 mg of MPA and 5 mg of estradiol cypionate per 0.5 mL. In women who receive repeated administration of Lunelle, peak estradiol levels occur approximately 2 days after the third injection and are 247 pg/mL (similar to peak ovulatory levels). The estradiol level returns to baseline 14 days after the last injection (100 pg/mL); the drop in estradiol is associated with menstrual bleeding (2-3 weeks after the last injection). Peak MPA levels (2.17 ng/mL) occur 3½ days after the third monthly injection. The mean MPA level is 1.25 ng/mL. The level at day 28 of the cycle is 0.44 to 0.47 ng/mL (level needed for contraceptive effect is 0.1-0.2 ng/mL). The earliest return of ovulation seen in women with multiple injections has been 60 days after the last dose. The mechanism of action, risks, and contraindications are similar to those of combined oral contraceptives.

Lunelle is administered intramuscularly in the buttock or deltoid every month. The first injection should be given within the first 5 days of the menstrual cycle (see Table 13–10). Even though

pharmacokinetic analysis reveals a delay in ovulation, the manufacturer recommends a 5-day grace period. The failure rate is 0.1 per 100 woman-years. Neither body weight nor use of concomitant drugs appears to affect the efficacy. Although this contraceptive has the advantages of the oral contraceptives and is associated with better compliance, the continuation rate is only 55%. This may be due to its side-effect profile, which is similar to that of the combined oral contraceptives with the addition of monthly injections.

There are limited data on potential risks. The risk potential is probably similar to that of the combined oral contraceptives, with a potentially lower incidence of deep vein thrombosis secondary to the absence of the first-pass effect. On discontinuation of Lunelle, achieving pregnancy may be delayed for as long as to 3 to 10 months after the last injection.

Lunelle was approved for use in the United States in 2000. However, in October 2002, Lunelle was recalled from the market, due to plant manufacturing problems. Alternatives to Lunelle that are administered outside the United States are: Mesigyna, Perlutal, Yectames, and Chinese Injectable No 1.

Subdermal Implants

The Norplant package consists of six capsules (34 mm in length, 2.4 mm in diameter), with each capsule providing 36 mg of levonorgestrel (total 216 mg). The target population is women who have contraindications to or adverse side effects from estrogen, women who are postpartum or breastfeeding, and adolescent mothers. This method provides long-term continuous contraception (approved for 5 years) that is rapidly reversible. The advantages, side effects, risks, and contraindications are similar to those of oral progestins. The major disadvantage—not present with use of oral progestins—is the surgical insertion and removal of the rods. A newer system, Norplant II, contains two rods (4 cm in length, 3.4 cm in diameter) and releases 50 μg/d of norgestrel (approved for 3 years). The two-rod system has the same mechanism of action and side-effect profile as its predecessor. However, the rods are much easier and faster to insert and remove than the capsules.

Within the first 24 hours, serum concentrations of levonorgestrel are 0.4 to 0.5 ng/mL. The capsules release 85 μg of levonorgestrel per 24 hours for the first year (equivalent to the daily dose of progestin-only pills) and then 50 μg for the remaining 5 years. The mean serum levels of progestin after the first 6 months are 0.25 to 0.6 ng/mL, slightly decreased at 5 years to 0.17 to 0.35 ng/mL. A levonorgestrel concentration below 0.2 ng/mL is associated with increased pregnancy rates. The site of implantation (leg, forearm, and arm) does not affect circulating progestin levels. Even though progestin levels are sufficient to prevent ovulation within the first 24 hours, the manufacturer recommends use of a backup method for 3 days after insertion. On removal, the progestin levels rapidly decline and undetectable serum levels are achieved after 96 hours. As a result, most women ovulate within 1 month after removal of the implants.

Norplant provides contraception in several ways. In the first 2 years, the levonorgestrel concentration is high enough to suppress the LH surge—most likely at the hypothalamic level—and

thereby inhibits ovulation. However, given the low concentrations of progestin, there is no real effect on FSH. The estradiol levels approximate those in ovulatory women. In addition, there are irregular serum peaks (often prolonged) and declines in serum estrogen levels that may contribute to erratic bleeding. By 5 years, more than 50% of the cycles are ovulatory. However, ovulatory cycles while using Norplant have been associated with luteal phase insufficiency. Other mechanisms of contraception are similar to oral progestins and include thickening of the cervical mucus, alterations of the endometrium, and changes in tubal and uterine motility.

The failure rate is 0.2 to 2.1 failures per 100 woman-years (0.9 per 100 users). Like oral progestins, body weight affects circulating levels and may result in more failures in the fourth or fifth year of use. Similar to oral progestins, the incidence of ectopic pregnancy among failures is increased to 20% (overall incidence is 0.28-1.3 per 1000 woman-years). The continuation rate (discontinuation rate of 10%-15% per year) is age-dependent and ranges from 33% to 78%. Menstrual disturbances are the most frequent side effect; they approach 40% to 80%, especially in the first 2 years. Although the incidence of abnormal uterine bleeding is similar to that of Depo-Provera, a significant difference between these methods is that Norplant provides only a 10% amenorrhea rate at 5 years. Other reported side effects include headache (30% indication for removal) and possibly weight gain, mood changes, anxiety, and depression—as well as ovarian cyst formation (eightfold increase), breast tenderness, acne, galactorrhea (if insertion occurs on discontinuation of lactation), possible hair loss, and pain or other adverse reactions at the insertion site (0.8% of cases at discontinuation).

Transdermal Patch

The transdermal patch (Ortho Evra) is another approach to contraception. The thin 20-cm^2 patch is composed of a protective layer, a middle (medicated) layer, and a release liner that is removed prior to application. The system delivers 150 μg of norelgestromin (active metabolite of norgestimate) and 20 μg of ethinyl estradiol per day to the systemic circulation. The target population is similar to that described earlier for Lunelle. One advantage of this system over Lunelle is that there are no monthly injections, and as a result, there is greater autonomy for the patient. The patch is applied once a week for 3 consecutive weeks, followed by a patch-free week for monthly withdrawal bleeding. The patch should be changed on the same day each week. The mechanism of action, contraindications, and side effects are similar to what has been described in the section on oral contraceptives.

With use of the transdermal patch, the peak ethinyl estradiol and norelgestromin levels are 50 to 60 pg/mL and 0.7 to 0.8 ng/mL, respectively. Because of this unique delivery system, hormone levels achieve a steady state condition throughout the cycle (see later and Figure 13–24). After the seventh day of application, there are adequate hormone levels to inhibit ovulation for 2 more days. With each consecutive patch, there is minimal accumulation of norelgestromin or ethinyl estradiol. The amount of hormone delivered is not affected by the environment, activity, or site of
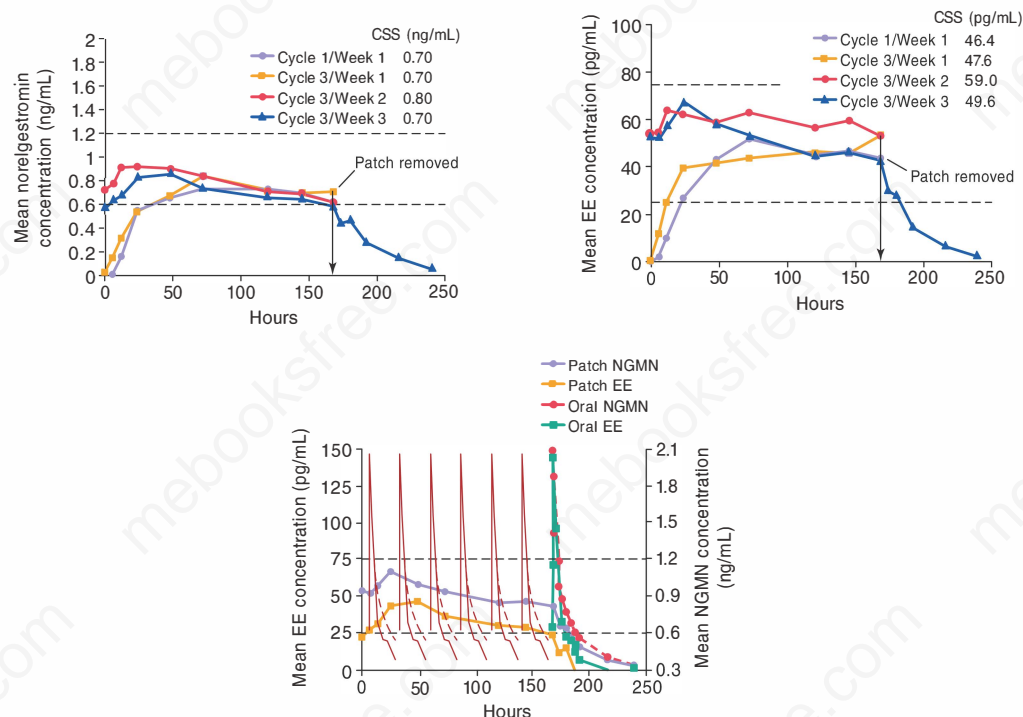
FIGURE 13–24 Comparative serum steroid (CSS) levels of norelgestromin (NGMN) and ethinyl estradiol (EE) following patch administration. (Reproduced with permission from Abrams LS et al. Multiple-dose pharmacokinetics of a contraceptive patch in healthy women participants. *Contraception*. 2001 Nov;64(5):287-294.)

application (abdomen, buttock, arm, torso). The adhesive is very reliable in a variety of conditions, including exercise, swimming, humidity, saunas, and bathing. Complete detachment occurs in 1.8% of cases and partial detachment in 2.9% of cases.

The failure rate is 0.7 per 100 woman-years under conditions of perfect use. Body weight has not been shown to affect the efficacy. The compliance with perfect use ranges from 88.1% to 91% among all age groups. This is significantly different from what is achieved with oral contraceptives (67%-85%), especially with women under 20 years of age. The side-effect profile is similar to that of oral contraceptives, except that there is slightly more breakthrough bleeding with the transdermal patch in the first 1 to 2 months (up to 12.2% vs 8.1%) and less breast tenderness (6.1% vs 18.8%). The incidence of skin reaction was 17.4%, characterized as mild in 92%, resulting in discontinuation in under 2%. The risks and contraindications are similar to that of combined oral contraceptives.

Vaginal Rings

Since the early 1900s, it has been recognized that the vagina is a place where steroids can be rapidly absorbed into the circulation. A study in the 1960s revealed that silicone rubber pessaries containing sex steroids would release the drug at a continuous rate. These studies led to the development of contraceptive vaginal rings.

As with oral contraceptives, there are combination and progestin-only formulations. Several progestin-only rings have been introduced since the 1970s. However, they were associated with significant menstrual disturbances. More recently, combination types have been developed. The most recent (2002) FDA-approved vaginal ring is a combination type called the NuvaRing.

The NuvaRing is made of ethylene vinyl acetate that provides 0.015 mg of ethinyl estradiol and 0.120 mg of etonogestrel per day. Maximum serum concentrations are achieved within 1 week after placement. The ring is designed to be used for 21 days and then removed for 1 full week to permit withdrawal bleeding. This device is capable of inhibiting ovulation within 3 days after insertion. After removal, the time to ovulation is 19 days. The mechanism of action, contraindications, and risks are similar to those of oral contraceptives. However, when assessing systemic exposure, use of the vaginal ring allows for 50% of the total exposure to ethinyl estradiol (15 μg in the ring compared with a 30-μg ethinyl estradiol-containing oral contraceptive).

The failure rate is similar to that reported with oral contraceptives. The continuation rate is 85.6% to 90%. Irregular bleeding is minimal (5.5%), and overall, the device is well tolerated, with an associated 2.5% discontinuation rate. Side effects are similar to those of oral contraceptives, but cycle control appears to be improved. The reported incidence of vaginal discharge is 23% versus 14.5% with oral contraceptives. The ring does not appear to interfere with intercourse (1%-2% of partners reported discomfort); however, the device can be removed for 2 to 3 hours during intercourse without changing efficacy.

Intrauterine Devices

IUDs are another modality of contraception that has been used clinically since the 1960s. Historically, these devices were made of plastic (polyethylene) impregnated with barium sulfate to make them radiopaque. Several other devices were subsequently developed, including the Dalkon Shield. After the introduction of the Dalkon Shield, an increase in pelvic infections was observed secondary to its multifilament tail. Furthermore, tubal infertility and septic abortions were increasing, and massive litigation occurred as a result. Consequently, even though the modern IUD has negligible associated risk, the use of IUDs in the United States is minimal—less than 1% of married women.

Currently, two types of IUDs are used in the United States: the copper- and the hormone-containing devices. The most recent FDA-approved intrauterine system contains levonorgestrel (Mirena) and is approved for 5 years of use. Several studies have demonstrated that these devices are unlike the Dalkon Shield and are very safe and efficacious. The target population is women who desire highly effective contraception that is long-term and rapidly reversible.

The copper (TCu-380A) IUD is a T-shaped device. The mechanism of action is mostly spermicidal due to the sterile inflammatory reaction that is created secondary to a foreign body in the uterus. The abundance of white blood cells that are present as a result kills the spermatozoa by phagocytosis. The amount of dissolution of copper is less than the daily amount ingested in the diet. However, with release of copper, salts are created that alter the endometrium and cervical mucus. Sperm transport is significantly impaired, limiting access to the oviducts.

There are two hormone-containing IUDs: the progesterone-releasing device (Progestasert) and the levonorgestrel-releasing device (Mirena). The Progestasert contains progesterone, which is released at a rate of 65 mg/d (approved for 1 year). This diffuses into the endometrial cavity, resulting in decidualization and atrophy of the endometrium. Serum progesterone levels do not change with the use of Progestasert. The main mechanism of action is to impair implantation. Mirena contains 52 mg of levonorgestrel, which is gradually released at a rate of 20 μg/d (approved for 5 years). Unlike Progestasert, systemic absorption of levonorgestrel inhibits ovulation about half the time. Although women may continue to have cyclic menses, over 40% have impaired follicular growth, with up to 23% developing luteinized unruptured follicles. Other mechanisms of action are similar to those described for Progestasert and the progestin-only pills. Mirena has the added advantage of significantly decreasing menstrual flow and has been used to treat menorrhagia.

The IUD should be placed within 7 days after onset of the menstrual cycle or at any time postpartum. Protection begins immediately after insertion. The failure rates after the first year of use are 0.5% to 0.8% for the TCu-380A, 1.3% to 1.6% for Progestasert, and 0.1% to 0.2% for Mirena. The expulsion rate is approximately 10%.

If a woman becomes pregnant with an IUD in place, the incidence of an ectopic pregnancy is 4.5% to 25%. The incidence of ectopic pregnancies with IUDs varies depending on the type of device. With Progestasert, the ectopic rate is slightly higher (6.80 per 1000 woman-years), most likely because its mechanism of action is limited to inhibiting implantation in the endometrium—in contrast to the copper or levonorgestrel IUD

(0.2-0.4 per 1000 woman-years), both of which also interfere with conception.

The continuation rate range for the current IUDs is 40% to 66.2% (Mirena). The side effects of the copper IUD include dysmenorrhea and menorrhagia. The most common adverse effect associated with the hormone-containing devices is erratic, albeit significantly less, bleeding. In fact, 40% of women experienced amenorrhea at 6 months and 50% at 12 months. The incidence of spotting in the first 6 months was 25% but decreased to 11% after 2 years. Other reported side effects of levonorgestrel include depression, headaches, and acne. There is a tendency to develop ovarian cysts early after insertion with the levonorgestrel-containing device that resolves after 4 months of use.

The nominal risks associated with IUD use include pelvic infection (within 1 month after insertion), lost IUD (ie, perforation into the abdominal cavity; 1:3000), and miscarriage. There is no association between IUD use and uterine or cervical cancer. Contraindications to IUD use are active genital infection and unexplained bleeding.

EMERGENCY CONTRACEPTION

Postcoital contraception is a method that may be used by a woman who believes her contraceptive method has failed or who has had unprotected intercourse and feels that she may be at risk for an unintended pregnancy. The first study to evaluate the efficacy of emergency contraception with hormones was in 1963. Subsequently, several studies have been performed with various contraceptives that paved the way for more widespread use. In 1997, the FDA approved the use of high-dose oral contraceptives for postcoital contraception. Since then, pharmaceutical companies have marketed specific packaging for the use of emergency contraception. Other methods, including mifepristone (RU-486) and the IUD have also been effective for postcoital contraception.

Similar to the formulas of oral contraception, there are both combination and progestin-only types of emergency contraception. The combination method (Yuzpe regimen) entails administration of two doses of two tablets (Ovral: 50 µg ethinyl estradiol, 0.5 mg norgestrel) 12 hours apart (total: 200 µg ethinyl estradiol, 2 mg norgestrel). Other oral contraceptives may be used with adjustment in the number of pills for equivalence (ie, two doses of four tablets: any second-generation oral contraceptive 12 hours apart). The specific medication (Preven) that is FDA-approved and marketed for postcoital contraception contains two doses of four tablets (ethinyl estradiol 50 µg, levonorgestrel 0.25 mg) 12 hours apart. The progestin-only method involves two doses of 10 pills (Ovrette 0.075 mg) 12 hours apart (Plan A). The marketed form (Plan B) contains two doses of one tablet (levonorgestrel 0.75 mg) 12 hours apart.

Following a single oral dose of 0.75 mg of levonorgestrel, the serum concentration peak (5-10 ng/mL) was at 2 hours with a rapid decline during the first 24 hours. The mechanism of action is uncertain, but levonorgestrel most likely inhibits ovulation and alters the endometrium to prevent implantation. Studies have shown decreased sperm recovery from the uterine cavity, possibly

due to thickened cervical mucus, or the alkalinization of the intrauterine environment. Others have shown that changes in other factors, such as integrins, can alter endometrial receptivity. The mode of action likely depends on the timing of intercourse relative to ovulation and to the administration of emergency contraception.

Maximum efficacy is achieved if the first dose is administered within 72 hours after intercourse and repeated in 12 hours. Combination formulas have a failure rate of 2% to 3%, and progestin-only preparations have a failure rate of 1%. Emergency contraception effectively reduces the rate of unintended pregnancies from 8% to 2%, a 75% reduction. However, with increasing time since unprotected intercourse, the efficacy changes from 0.4% to 1.2% to 2.7% for the first, second, or third 24-hour period after unprotected intercourse. For maximum efficacy, emergency contraception may be prescribed in advance so women will already have the correct dosing. No increase in risk-taking behavior has been noted with this strategy.

Significant nausea or emesis (51.7%) is associated with use of emergency contraception, although it is substantially less with progestin-only formulations. An antiemetic should be administered 1 hour before each treatment. If a patient vomits within 1 hour after ingestion, additional pills need to be administered.

Contraindications to emergency contraception with the combination regimen are possibly the same as those described for oral contraceptives; for progestin-only pills, there are no contraindications. Emergency contraception should be an optional function of the rape management protocol.

REFERENCES

Embryology and Anatomy

Ahmed SF, Hughes IA. The genetics of male under-masculinization. *Clin Endocrinol (Oxf)*. 2002;56:1. [PMID: 11849240]

McElreavey K, Fellous M. Sex determination and the Y chromosome. *Am J Med Genet*. 1999;89:76. [PMID: 10727993]

Oocyte Development

Braw-Tal R. The initiation of follicle growth: the oocyte or the somatic cells? *Mol Cell Endocrinol*. 2002;18:18. [PMID: 11988306]

Erickson GF. The ovary: basic principles and concepts. In: Felig P, Frohman LA, eds. *Endocrinology and Metabolism*. 4th ed. New York, NY: McGraw-Hill; 2001.

Russell DL, Robker RL. Molecular mechanisms of ovulation: co-ordination through the cumulus complex. *Hum Reprod Update*. 2007;13:289. [PMID: 17242016]

Teixeira J, Maheswaran S, Donahoe PK. Müllerian inhibiting substance: an instructive developmental hormone with diagnostic and possible therapeutic applications. *Endocr Rev*. 2001;22:657. [PMID: 11588147]

Zhang M, Su Y-Q, Sugiura K, Xia G, Eppig JJ. Granulosa cell ligand NPPC and its receptor NPR2 maintain meiotic arrest in mouse oocytes. *Science*. 2010;330:366. [PMID: 20947764]

Ovarian Steroidogenesis

Bulun SE, Adashi EY. The physiology and pathology of the female reproductive axis. In: Larsen et al, ed. *Williams Textbook of Endocrinology*. 10th ed. Philadelphia, PA: WB Saunders; 2003.

Strauss JF, Hsueh AJ. Ovarian hormone synthesis. In: DeGroot LJ et al, ed. *Endocrinology*. 4th ed. Philadelphia, PA: WB Saunders; 2001.

Menstrual Cycle

Chabbert-Buffet N, Bouchard P. The normal human menstrual cycle. *Rev Endocr Metab Disord*. 2002;3:173. [PMID: 12215712]

McGee EA, Hsueh AJW. Initial and cyclic recruitment of ovarian follicles. *Endocr Rev*. 2000;21:200. [PMID: 10782364]

Roseweir AK, Millar RP. The role of kisspeptin in the control of gonadotrophin secretion. *Hum Reprod Update*. 2009:15:203. [PMID: 19109311]

Amenorrhea

Fourman LT, Fazeli PK. Neuroendocrine causes of amenorrhea—an update. *J Clin Endocrinol Metab*. 2015;100:812. [PMID: 25581597]

Marcus MD, Loucks TL, Berga SL. Psychological correlates of functional hypothalamic amenorrhea. *Fertil Steril*. 2001;76:310. [PMID: 11476768]

Marshall JC, Eagleson CA, McCartney CR. Hypothalamic dysfunction. *Mol Cell Endocrinol*. 2001;183:29. [PMID: 11604221]

Ovarian Failure

Bakalov VK, Vanderhoof VH, Bondy CA, Nelson LM. Adrenal antibodies detect asymptomatic auto-immune adrenal insufficiency in young women with spontaneous premature ovarian failure. *Hum Reprod*. 2002;17:2096. [PMID: 12151443]

Christin-Maitre S, Vasseur C, Portnoi M-F, Bouchard P. Genes and premature ovarian failure. *Mol Cell Endocrinol*. 1998;145:75. [PMID: 9933102]

Davison RM, Davis CJ, Conway GS. The X chromosome and ovarian failure. *Clin Endocrinol*. 1999;51:573. [PMID: 10619970]

Layman LC. Human gene mutations causing infertility. *J Med Genet*. 2002;3:153. [PMID: 11897813]

Anovulation Due to Polycystic Ovarian Syndrome

Abbott DH, Dumesic DA, Franks S. Developmental origin of polycystic ovary syndrome—a hypothesis. *J Endocrinol*. 2002;174:1. [PMID: 12098657]

Azziz R, Carmina E, Dewailly D, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil Steril*. 2009;91:456. [PMID: 18950759]

Ibanez L, Dimartino-Nardi J, Potau N, Saenger P. Premature adrenarche—normal variant or forerunner of adult disease? *Endocr Rev*. 2000;21:671. [PMID: 11133068]

Poretsky L, Cataldo NA, Rosenwaks Z, Giudice LC. The insulin-related ovarian regulatory system in health and disease. *Endocr Rev*. 1999;20:535. [PMID: 10453357]

Anovulation Due to Adrenal Disorders

Betterle C, Dal Pra C, Mantero F, Zanchetta R. Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction. *Endocr Rev*. 2002;23:327. [PMID: 12050123]

White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr Rev*. 2000;21:245. [PMID: 10857554]

Müllerian Anomalies

Homer HA, Li T-C, Cooke ID. The septate uterus: a review of management and reproductive outcome. *Fertil Steril*. 2000;73:1. [PMID: 10632403]

Raga F, Bauset C, Remohi J, et al. Reproductive impact of congenital Müllerian anomalies. *Hum Reprod*. 1997;12:2277. [PMID: 9402295]

Menopause: General

Couzinet B, Meduri G, Lecce MG, et al. The postmenopausal ovary is not a major androgen-producing gland. *J Clin Endocrinol Metab*. 2001;86:5060. [PMID: 11600585]

Faddy MJ, Gosden RG, Gougeon A, et al. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. *Hum Reprod*. 1992;7:1342. [PMID: 1291557]

Hansen KR, Knowlton NS, Thyer AC, et al. A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. *Hum Reprod*. 2008;23:699. [PMID: 18192670]

Harlow SD, Gass M, Lobo R, et al. For the STRAW + 10 Collaborative Group. Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. *Menopause*. 2012;19:387. [PMID: 22338612]

Nair GV, Herrington DM. The ERA trial: findings and implications for the future. *Climacteric*. 2000;3:227. [PMID: 11910581]

Randolph JF Jr, Sowers M, Bondarenko IV, et al. Change in estradiol and follicle-stimulating hormone across the early menopausal transition: effects of ethnicity and age. *J Clin Endocrinol Metab*. 2004;89:1555. [PMID: 15070912]

Reddy SY, Warner H, Guttuso T, et al. Gabapentin, estrogen, and placebo for treating hot flushes: a randomized controlled trial. *Obstet Gynecol*. 2006;108:41. [PMID: 16816054]

Rosen M, Sternfeld B, Schuh-Huerta SM, et al. Antral follicle count: absence of significant midlife decline. *Fertil Steril*. 2010;94:2182. [PMID: 20149366]

Sowers MR, Eyvazzadeh AD, McConnell D, et al. Anti-Mullerian hormone and inhibin B in the definition of ovarian aging and the menopause transition. *J Clin Endocrinol Metab*. 2008;93:3478. [PMID: 18593767]

Weiss G, Skurnick JH, Goldsmith LT, et al. Menopause and hypothalamic-pituitary sensitivity to estrogen. *JAMA*. 2004;292:2991. [PMID: 15613667]

Menopause and Cardiovascular Disease

Rossouw JE, Anderson GL, Limacher M, et al. Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*. 2004;291:1701. [PMID: 15082697]

U.S. Preventive Services Task Force. Postmenopausal hormone replacement therapy for primary prevention of chronic conditions: recommendations and rationale. *Ann Intern Med*. 2002;137:834. [PMID: 12435221]

Menopause: Osteoporosis

O'Connor KM. Evaluation and treatment of osteoporosis. *Med Clin North Am*. 2016;100:807. [PMID: 27235616]

Menopause: Selective Estrogen Receptor Modulators

Miller CP. SERMs: evolutionary chemistry, revolutionary biology. *Curr Phar Des*. 2002;8:2089. [PMID: 12171520]

Hormone Replacement and Alzheimer Disease

Fillit HM. The role of hormone replacement therapy in the prevention of Alzheimer's disease. *Arch Intern Med*. 2002;162:1934. [PMID: 12230415]

Infertility

Cedars MI. Infertility in obstetrics and gynecology: a longitudinal approach. In: Moore TR et al, eds. *Gynecology and Obstetrics: A Longitudinal Approach*. New York, NY: Churchill Livingstone; 1993.

Legro RS, Brzyski RG, Diamond MP, et al. Letrozole versus clomiphene for infertility in the polycystic ovary syndrome. *N Engl J Med*. 2014;371:119. [PMID: 25006718]

Te Velde ER, Dorland M, Broekmans FJ. Age at menopause as a marker of reproductive ageing. *Maturitas*. 1998;30:119. [PMID: 9871906]

Contraception

Abma JC, Chandra A, Mosher WD, et al. Fertility, family planning, and women's health: new data from the 1995 National Survey of Family Growth, Report No. 19, Series 23, 1997.

Archer DF. New contraceptive options. *Clin Obstet Gynecol*. 2001;44:122. [PMID: 11219241]

Brenner PF, Mishell DR Jr, Stanczyk FZ, Goebelsmann U. Serum levels of d-norgestrel, luteinizing hormone, follicle-stimulating hormone, estradiol, and progesterone in women during and following ingestion of combination

oral contraceptives containing dl-norgestrel. *Am J Obstet Gynecol*. 1977;129:133. [PMID: 900174]

Burkman RT, Collins JA, Shulman LP, Williams JK. Current perspectives on oral contraceptive use. *Am J Obstet Gynecol*. 2001;185(2 suppl):S4. [PMID: 11521117]

Croxatto HB, Devoto L, Durand M, et al. Mechanism of action of hormonal preparations used for emergency contraception: a review of the literature. *Contraception*. 2001;63:111. [PMID: 11368982]

Kaunitz AM. Injectable long-acting contraceptives. *Clin Obstet Gynecol*. 2001;44:73. [PMID: 11219248]

Marchbanks PA, McDonald JA, Wilson HG, et al. Oral contraceptives and the risk of breast cancer. *N Engl J Med*. 2002;346:2025. [PMID: 12087137]

McCann MF, Potter LS. Progestin-only oral contraception: a comprehensive review. *Contraception*. 1994;50(6 suppl):S1. [PMID: 10226677]

Ortiz A, Hirol M, Stanczyk FZ, et al. Serum medroxyprogesterone acetate (MPA) and ovarian function following intramuscular injection of depo-MPA. *J Clin Endocrinol Metab*. 1977;44:32. [PMID: 833262]

Rahimy MH, Ryan KK, Hopkins NK. Lunelle monthly contraceptive injection (medroxyprogesterone acetate and estradiol cypionate injectable suspension): steady-state pharmacokinetics of MPA and E2 in surgically sterile women. *Contraception*. 1999;60;209. [PMID: 10640167]

This page intentionally left blank

Disorders of Sex Development

Rodolfo A. Rey, MD, PhD, Christopher P. Houk, MD,
Selma Witchel, MD, and Peter A. Lee, MD, PhD

17-OHP	17-hydroxyprogesterone	**CYP19A1**	Cytochrome P450, family 19, subfamily A, polypeptide 1
21-OHD	21-hydroxylase deficiency		
ABS	Antley-Bixler syndrome	**CYP21A1P**	Cytochrome P450, family 21, subfamily A, polypeptide 1 pseudogene
ACTH	Adrenocorticotropin		
AGD	Asymmetric gonadal differentiation (synonym Mixed Gonadal Dysgenesis, MGD)	**CYP21A2**	Cytochrome P450, family 21, subfamily A, polypeptide 2
		CYP26B1	Cytochrome P450, family 26, subfamily B, polypeptide 1
AIS	Androgen insensitivity syndrome		
AKR1C	Aldo-ketoreductase	**CYP51A1**	Cytochrome P450, family 51, subfamily A, polypeptide 1
ALDH	Aldehyde dehydrogenase (synonym: Retinaldehyde Dehydrogenase, RALDH)	**DACH**	Dachshund
		DAX1	DSS (dosage-sensitive sex reversal)-associated gene 1
AMH	Anti-Müllerian hormone (synonym: Müllerian Inhibiting Substance, MIS)		
		DELFIA	Dissociation-enhanced lanthanide fluoroimmunoassay
AMHR	AMH receptor		
AR	Androgen receptor	**DHCR**	Dehydrocholesterol reductase
ARE	Androgen response element	**DHEA**	Dehydroepiandrosterone
ATRX	Alpha-thalassemia, mental retardation gene on the X chromosome	**DHH**	Desert hedgehog
		DHT	Dihydrotestosterone
AXIN	Axis inhibitor	**DM**	Doublesex- and MAB3
BMP	Bone morphogenetic protein	**DMRT**	Doublesex- and MAB3-related transcription factor
C4	Complement component 4		
CAH	Congenital adrenal hyperplasia	**DNA**	Deoxyribonucleic acid
CAIS	Complete androgen insensitivity syndrome	**DOC**	Deoxycorticosterone
		DSD	Disorder of sex development
CBX2	Chromobox homolog gene 2	**EMX**	Empty spiracles gene
CFTR	Cystic fibrosis transmembrane conductance regulator	**FGF**	Fibroblast growth factor
		FGFR	Fibroblast growth factor receptor
CGH	Comparative genomic hybridization	**FISH**	Fluorescent in situ hybridization
CGPR	Calcitonin gene-related peptide	**FOG**	Friend of GATA
CITED	CBP/p300-interacting transactivator	**FOXL2**	Forkhead transcription factor 2
C-KIT	Kit oncogene	**FSH**	Follicle-stimulating hormone
CNS	Central nervous system	**FST**	Follistatin
CNV	Copy-number variation	**GADD45G**	Growth arrest- and DNA damage-inducible gene 45 gamma
CYP11B1	Cytochrome P450, family 21, subfamily B, polypeptide 1		
		GATA4	GATA-binding protein 4
CYP17A1	Cytochrome P450, family 17, subfamily A, polypeptide 1		

GC-MS	Gas chromatography-mass spectrometry	**PAIS**	Partial androgen insensitivity syndrome
GSK3β	Glycogen synthase kinase 3-beta	**PAR**	Pseudoautosomal region
GWAS	Genome-wide association studies	**PAX**	Paired box gene
hCG	Human chorionic gonadotropin	**PCOS**	Polycystic ovary syndrome
HMG	High mobility group	**PCR**	Polymerase chain reaction
HOX	Homeobox	**PDGF**	Platelet-derived growth factor
HRQOL	Health-related quality of life	**PGD**	6-phosphogluconate dehydrogenase
HSD	Hydroxysteroid dehydrogenase	**PKD**	Polycystic kidney disease
HSD3B	3-beta-hydroxysteroid dehydrogenase	**PMDS**	Persistent Müllerian duct syndrome
IGF1	Insulin-like growth factor 1	**POMC**	Proopiomelanocortin
IGFBP3	Insulin-like growth factor binding protein 3	**POR**	Cytochrome P450 oxidoreductase
IM	Intramuscular	**PRDM**	PR domain-containing protein (synonym: B lymphocyte-induced maturation protein, BLIMP)
INSL3	Insulin-like factor 3		
IU	International unit	**PTEN**	Phosphatase and tensin
LC-MS/MS	Liquid chromatography-mass spectrometry in tandem	**QoL**	Quality of life
		RNA	Ribonucleic acid
LEF	Lymphoid enhancer-binding factor 1	**RP**	Retinitis pigmentosa
LGR4	Leucine-rich repeat-containing G protein-coupled receptor	**RSPO1**	Roof plate-specific spondin 1
		S compound	11-deoxycortisol
LH/CGR	Luteinizing hormone/chorionic gonadotropin receptor	**SERKAL**	Sex reversal with dysgenesis of kidneys, adrenals and lungs
LH	Luteinizing hormone	**SF1**	Steroidogenic factor 1 (synonym: Adrenal 4 binding protein, Ad4BP)
LHX	LIM homeobox gene		
LHX9	LIM homeobox gene 9	**SHOX**	Short stature homeobox
MAMLD1	Mastermind-like domain-containing protein 1	**SIX**	Sine oculis homeobox genes
		SLOS	Smith-Lemli-Opitz Syndrome
MAP3K4	Mitogen-activated protein kinase 4	**SNP**	Single-nucleotide polymorphism
MGD	Mixed gonadal dysgenesis (synonym Asymmetric gonadal differentiation, AGD)	**SOX**	SRY-related box
		SRD5A2	Steroid 5-alpha-reductase 2
		SRY	Sex-determining region on the Y chromosome
MKKS	McKusick-Kaufman syndrome		
MRKH	Mayer-Rokitansky-Küster-Hauser syndrome	**StAR**	Steroidogenic acute regulatory protein
		STRA8	Stimulated by retinoic acid gene 8
MURCS	Müllerian agenesis associated with urinary tract abnormalities, and even cervicothoracic somite dysplasia	**TART**	Testicular adrenal rest tumor
		TES	Testis specific
		TESCO	TES core element
NANOG	Homeobox transcription factor NaNog	**TGF**	Transforming growth factor
NCAH	Nonclassic congenital adrenal hyperplasia	**THF**	Tetrahydrocortisol
		THS	Tetrahydro-11-deoxycortisol
NR0B1	Nuclear receptor subfamily 0, group B, member 1	**TNX**	Tenascin
		TSPO	Translocator protein
NR3C1	Nuclear receptor subfamily 3, group C, member 1	**VANGL**	Van Gogh-like
		VEGF	Vascular endothelial growth factor
NR5A1	Nuclear receptor subfamily 5, group A, member 1	**WAGR**	Wilms tumor, aniridia, genitourinary anomalies and mental retardation
OART	Ovarian adrenal rest tumor		
OCT	Octamer binding transcription factor	**WES**	Whole exome sequencing
P450c17	Cytochrome P450 17-hydroxylase/17,20-lyase	**WGS**	Whole genome sequencing
		WNT	Wingless-type MMTV integration site
P450scc	Cytochrome P450 side-chain cleavage enzyme	**WT1**	Wilms tumor gene 1

NORMAL FETAL SEX DIFFERENTIATION

Fetal sexual development consists of three sequential stages (Figure 14–1): (1) an undifferentiated stage, characterized by the formation of gonadal and genital primordia, which are identical in both XX and XY embryos; (2) gonadal differentiation, where the undifferentiated gonads develop into ovaries or testes; and (3) sexual differentiation of the internal and external genitalia, which depends on the existence or absence of male hormone action. Other organs, for example, the brain, liver, or kidney, also undergo sexual differentiation, although their dimorphism is less evident. The understanding of these normal processes set the basis for the comprehension of the etiologies, pathophysiology, clinical presentation, and management of disorders of sex development (DSD). We will focus on those pathways known to have clinical relevance in humans, even if the knowledge derives from observations in experimental models.

THE UNDIFFERENTIATED STAGE

Between the fourth and the sixth weeks of embryonic development, that is, the sixth and eighth weeks from the last menstrual cycle, the anlagen of the gonads and reproductive ducts develop from the intermediate and lateral mesoderm. Endodermic derivatives involved in the formation of the cloaca and ectodermic structures give rise to the anlagen of future internal reproductive structures and of the external genitalia.

INITIAL FORMATION OF THE UROGENITAL RIDGES

During the fourth and fifth weeks, the intermediate mesoderm consists of paired cylindrical masses at each side of the midline, arranged along the posterior aspect of the embryo, laterally to the paraxial mesoderm and medially to the lateral plate mesoderm. The intermediate mesoderm thickens and protrudes into the coelomic cavity—the future abdominal cavity—to form the urogenital ridges, covered by the coelomic epithelium, a derivative of lateral mesoderm. Each urogenital ridge progressively becomes separated in two portions: a medial gonadal ridge and a lateral urinary ridge.

The urinary ridge can be divided into three regions: (1) at the cephalic end, the pronephros that rapidly regresses in the human embryo; (2) at the caudal end near the cloaca, the metanephros that gives rise to most of the kidney; (3) between them lies the mesonephros, which provides mesenchymal components to the gonadal ridge and also gives rise to the mesonephric or Wolffian ducts. The normal development of the Wolffian ducts is essential not only for the reproductive system, but also for the development of the kidneys.

Early embryonic developmental processes are driven by regulatory pathways involving ubiquitously expressed factors whose deficiency may be lethal or result in multiple malformations. For instance, in mice homeobox transcription factors, like LHX1, EMX2 or PAX3, and others like WT1 play a role in the stabilization of the intermediate mesoderm leading to the development of the urogenital ridges. Knockout models show absence or defects of all urogenital derivatives, that is, kidneys and ureters, gonads and internal genital ducts.
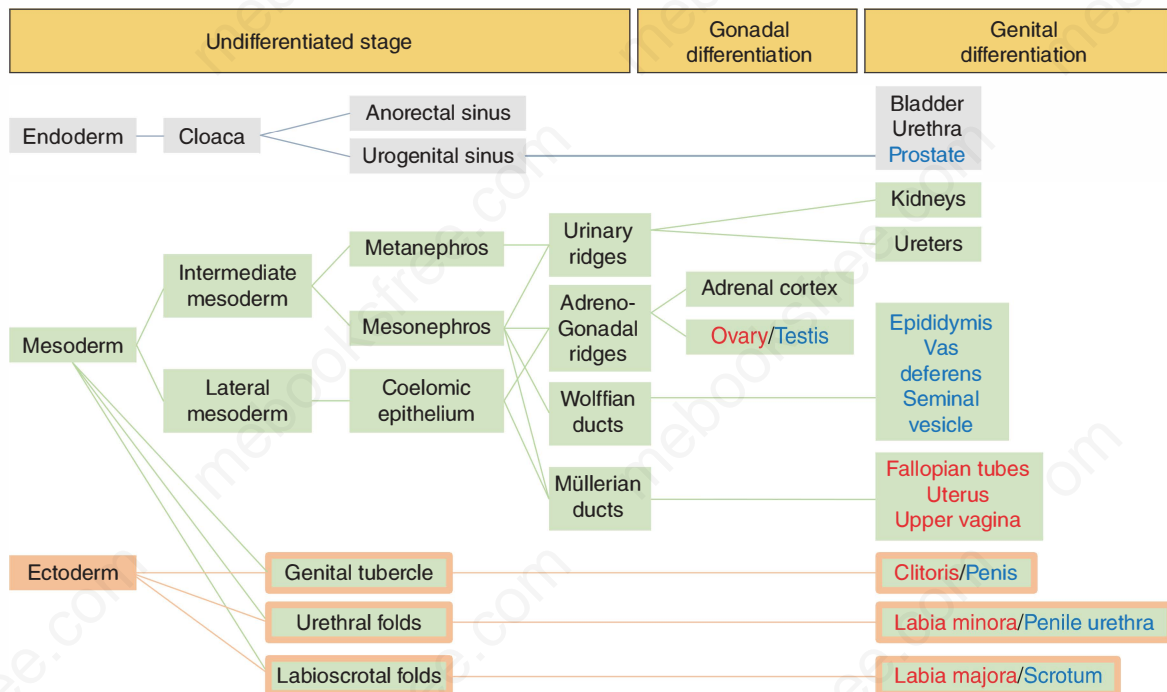
FIGURE 14–1 The three stages in fetal sex differentiation. Gray boxes indicate endodermal derivatives; green boxes, mesodermal derivatives; and green/red boxes, structures with both mesodermal and ectodermal derivatives.

THE BIPOTENTIAL GONADS

The gonads and the adrenal cortex differentiate from the adreno-gonadal primordia, the medial portions of the former urogenital ridges. Factors like SF1 and DAX1, respectively encoded by genes *NR5A1* and *NR0B1*, CBX2, LHX9, GATA4, and SIX1/4 are essential for the development of the adreno-gonadal primordia. Cells from the coelomic epithelium and the mesonephric mesoderm contribute to their formation. Progressively, two distinct cell populations separate to originate the adrenal primordium and the gonadal blastema at the beginning of the fifth embryonic week. During the fifth week, the gonadal blastema is colonized by primordial germ cells, of extra-gonadal origin. Several factors have been identified to have a role in the differentiation, migrations, and proliferation/apoptosis of primordial germ cells. Primordial germ cells derive from epiblast; factors involved in the specification of primordial germ cells include BMP2, BMP4, BMP8B, OCT3/4 (encoded by *POU5F1*), NANOG, SOX2, PRDM1, and PRDM14. At a very early stage of embryonic life, primordial germ cells move through the extra-embryonic region at the base of the allantois. In the fourth week, they can be identified in the yolk sac by their expression of PRDM1, alkaline phosphatase, OCT3/4 and the tyrosine kinase receptor C-KIT. Later, primordial germ cells become embedded in the wall of the hindgut, and migrate through the dorsal mesentery to reach the gonadal ridges. Chromatin modifications occur at this stage: genome-wide demethylation leads to erasure of genomic imprints or epigenetic marks. Like somatic cells of the gonadal ridges, germ cells are bipotential. We shall see that germ cells are essential for the development of the ovaries, but not the testes.

Until the end of the fifth week, XX and XY gonads are indistinguishable and bipotential, in that they can develop into ovaries or testes. Interestingly, not only are the optic and electron microscopic aspects identical but also the transcriptomes of the prospective XX and XY gonads are indistinguishable: factors that are later associated with testis fate, like SOX9 and FGF9, or with ovary fate, like DAX1, WNT4, and RSPO1, are expressed at similar levels in the bipotential XX and XY mouse gonads.

THE UNIPOTENTIAL INTERNAL DUCTS

The undifferentiated internal reproductive tract is similar in the XX and the XY embryo until the end of the sixth week and consists on each side of a set of two unipotential ducts: the Wolffian (mesonephric) giving rise to the male internal tract, and the Müllerian (paramesonephric) ducts, that are the anlagen of the female internal tract (Figure 14–2).
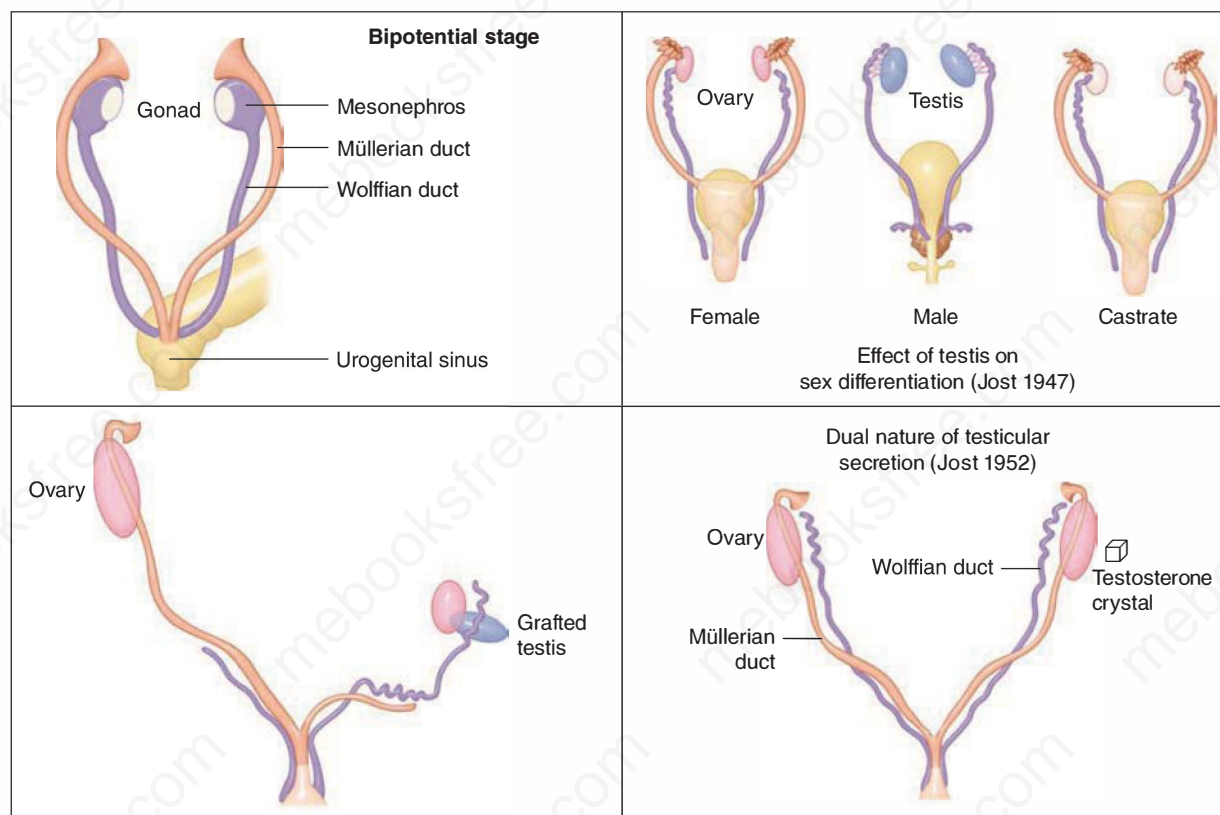
FIGURE 14–2 Jost's experiments revealing the determining role of the testes in fetal sex differentiation. In the male fetus, Wolffian ducts develop whereas Müllerian ducts regress. In the female fetus, the opposite occurs. Fetal castration, irrespective of genetic sex, results in female differentiation of the reproductive tract. (Reproduced with permission from Rey R, Josso N. Sexual differentiation. In: De Groot LJ, Beck-Peccoz P, Chrousos G, et al: Endotext [WWW.ENDOTEXT.ORG]. Published by MDTEXT.COM,INC, 2016. South DARTMOUTH,MA 02748.)

Wolffian Ducts

The Wolffian ducts originate in the intermediate mesoderm, dorsolaterally to the pronephros and laterally to somites 8 to 13, and grow caudally to end in the cloaca during the fifth embryonic week. A single ureteric bud evaginates from the Wolffian duct and grows dorsally, in response to inductive signals from metanephric mesenchyme. Interactions between the ureteric bud and the metanephros are essential for kidney development. Several factors have been identified as important for Wolffian duct formation. As already mentioned, members of the PAX family, WT1, EMX2, LIM1, and SIX1 are involved in early intermediate mesoderm stabilization. VANGL2, WNT5A, and PKD1 are necessary to Wolffian duct elongation, and PTEN, LGR4, and DICER, amongst others, play a role in Wolffian duct cell differentiation. Although the pathogenic mechanism is not fully understood, defects in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) are associated with congenital bilateral absence of Wolffian duct derivatives.

Müllerian Ducts

Müllerian ducts arise in the fifth embryonic week, as an invagination of the coelomic epithelium, laterally to the gonadal ridge. Their development can be divided into three phases: initiation, invagination and elongation. First, cells of the coelomic epithelium are specified to a Müllerian duct fate, they form placode-like thickening expressing LHX1, other members of the LIM family, DACH1 and DACH2, and anti-Müllerian hormone receptor type 2 (AMHR2), induced by WNT7A. In the second phase, Müllerian duct-specified cells extend caudally toward the Wolffian ducts. WNT4 expression in the mesonephric mesenchyme is necessary to begin invagination. Upon contact with the Wolffian duct, the elongation phase begins, consisting in proliferation and caudal migration of a group of cells at the caudal tip. Close contact with the Wolffian duct is necessary to Müllerian growth; indeed, the lack of Wolffian development leads to Müllerian truncation. Wolffian ducts do not contribute cells to the elongating Müllerian tip, but supply factors like WNT9B. In the eighth week, the growing tip of the Müllerian duct crosses the Wolffian duct ventrally and run in intimate contact with the contralateral Müllerian duct. These ducts fuse and give rise to the uterovaginal canal, which makes contact with the posterior wall of the urogenital sinus, causing an elevation, the Müllerian tubercle, flanked on both sides by the opening of the Wolffian ducts.

THE BIPOTENTIAL UROGENITAL SINUS AND EXTERNAL GENITALIA

The cloaca is the terminal portion of the hindgut, limited by the cloacal membrane. The epithelium of the cloaca is of endodermal origin, whereas the surrounding mesenchyme is mesodermal. In the fifth embryonic week (seventh week of amenorrhea), the urorectal septum divides the cloaca into the rectum dorsally and the urogenital sinus ventrally. The ventral portion of the cloacal

membrane becomes the urogenital membrane. Two mesodermal swellings develop laterally under the ectoderm of the genital membrane, the urethral folds and the labioscrotal swellings, while the genital tubercle emerges as a ventral and medial swelling. In the sixth week (8-15 mm embryo), the ostium of the urogenital sinus is surrounded by the labioscrotal swellings. Endodermal epithelial cells from the urogenital sinus invade the genital tubercle and form the urethral plate. After the corpora cavernosa and glans have differentiated, the ventral surface of the genital tubercle is depressed by a deep furrow, the urethral groove. The external genitalia remain undifferentiated up to approximately 9 weeks.

Early patterning of external genitalia is under the control of factors regulating the interactions between mesenchymal and epithelial tissues. As in the development of other body appendages, *BMP*, *FGF*, and *HOX* family genes play a predominant role during genital tubercle formation. SHH and β-catenin as well are key regulators of the mesenchymal genes involved in the patterning of the bipotential genital primordia.

At 10 weeks in both XX and XY fetuses, there is a vaginal primordium formed by the caudal tips of the Müllerian ducts, and medial and lateral outgrowths of the urogenital sinus, the sinovaginal bulbs, which fuse to form the vaginal cord or plate. When the vaginal plate regresses, the vaginal lumen is formed.

GONADAL DIFFERENTIATION

The differentiation of the gonadal ridge into a testis or an ovary is of paramount importance in the process of fetal sex differentiation. The pioneering experiments performed by Alfred Jost and colleagues more than 60 years ago clearly showed that, irrespective of their chromosomal constitution, when the gonads differentiate into testes, hormones produced drive internal and external genitalia along the male pathway. Conversely, when no testes are present, the genitalia follow the female pathway (see Figure 14–2). The ovaries have no influence on genital differentiation during intrauterine life. Recent progress in the understanding of molecular mechanisms indicates that the development of the undifferentiated gonadal ridge into a testis or an ovary requires a delicate dosage balance in the timing and levels of expression of several genes (Figure 14–3).

TESTICULAR DIFFERENTIATION

Testicular differentiation occurs by the end of the sixth embryonic week. Experimental evidence in mice indicates that the earliest hallmark of fetal testis development is the formation of the coelomic vessel in the XY gonadal ridge at its antimesonephric margin, that is, below the coelomic epithelium. A close spatial relationship exists between the gonadal ridge vascularization and testicular cord formation: endothelial cells migrating from the mesonephros to the coelomic zone of the differentiating gonad are essential for cord formation. Subsequently, somatic cells from the coelomic epithelium differentiate into Sertoli cells, aggregate within the
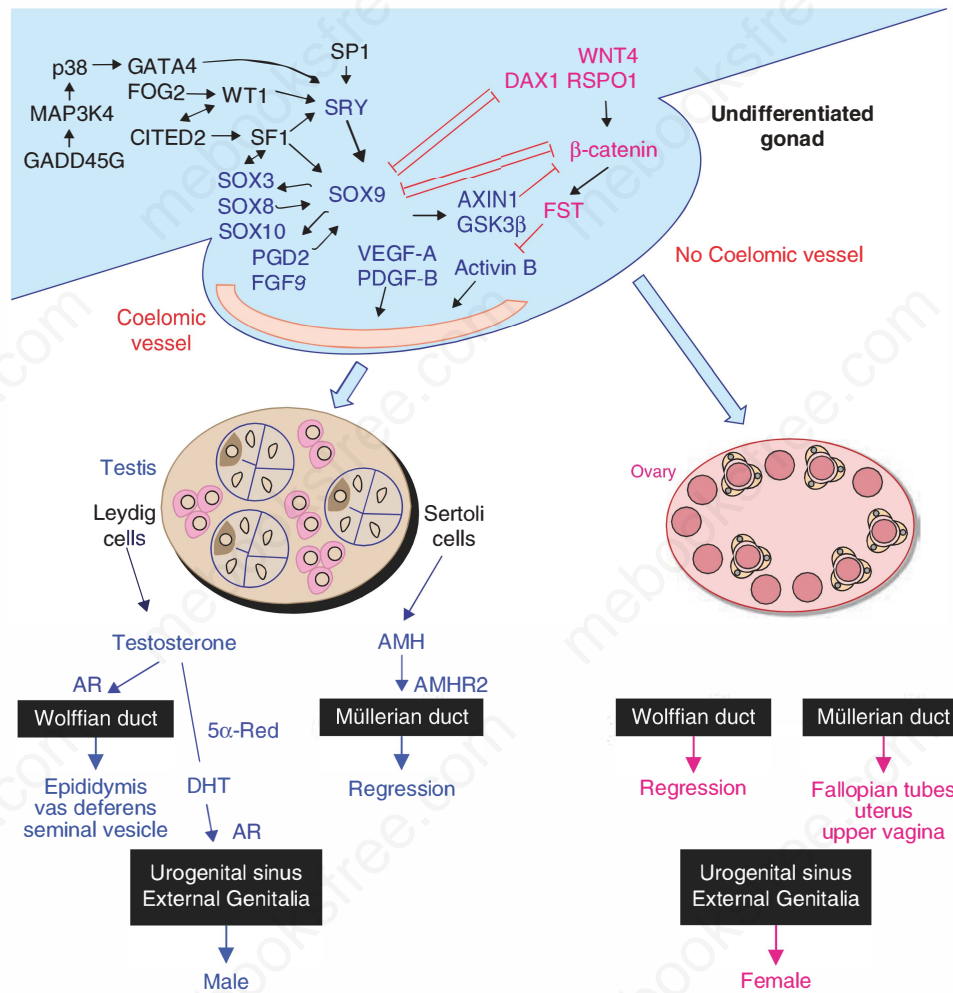
FIGURE 14–3 Gonadal and genital tract differentiation. Factors in black are involved in the development of the undifferentiated gonad; those in blue are essentially pro-testicular genes, whereas those in pink are mainly pro-ovarian or anti-testicular when overexpressed. When pro-testicular genes prevail, the coelomic vessel differentiates followed by testicular cord and Leydig cell development. Conversely, when no coelomic vessel develops, the gonad undergoes ovarian differentiation. The testes secrete two hormones involved in the virilization of the internal and external genitalia: anti-Müllerian hormone (AMH) by Sertoli cells, and testosterone by Leydig cells. In their absence, the genitalia follow the female pathway. 5α-red, 5α-reductase; AMH, anti-Müllerian hormone; AMHR2, AMH receptor type 2; AR, androgen receptor.

developing gonad and enclose germ cells to form the testicular cords, the precursors of the seminiferous tubules. Sertoli cells begin secreting anti-Müllerian hormone (AMH) as soon as they differentiate, that is, at the beginning of the seventh embryonic week. The primordial germ cells differentiate to gonocytes, which proliferate by mitosis but do not enter meiosis. Extracellular matrix components and mesenchymal cells of mesonephric origin fill the interstitial space between seminiferous cords and give rise to further blood vessels and differentiating Leydig cells, which begin secreting testosterone by the beginning of the eighth week.

OVARIAN DIFFERENTIATION

In the XX fetus, the gonad remains undifferentiated longer than in the XY fetus. Primordial germ cells differentiate to oogonia, which can be seen sparsely in the cortical zone during

the eighth week. The first unequivocal morphological sign of ovarian differentiation is meiotic entry of germ cells occurring in the tenth fetal week. Oocytes become surrounded by a single layer of follicular (or granulosa) cells to form primordial follicles. The earliest primary follicles appear at 15 weeks and the first Graafian follicles at 23 to 24 weeks. By the end of the seventh month of gestation, almost all germ cells have entered meiosis, proceeding to the diplotene stage, where they remain until gametogenesis is completed in adult life. Not all oocytes complete meiosis: from the 6 to 7 million ovarian follicles present at 25 fetal weeks, only 2 million persist at term. Most follicles undergo atresia.

The presence of germ cells is critical for the stabilization of the ovary, which is a major difference, as compared with testicular differentiation. While the somatic components of the fetal testis progress normally and produce testicular hormones in the absence of germ cells, ovarian follicles degenerate if germ cells are lost, and

the gonads acquire a fibrous aspect and are unable to synthesize ovarian hormones.

Granulosa cells, the female equivalent of Sertoli cells, enclose germ cells to form primordial follicles and regulate oocyte maturation. The source of granulosa cells is still debated: the coelomic epithelium, mesenchymal cells of the gonadal ridge and mesonephric cells form the adjacent rete ovarii. Theca cells surround the granulosa cells from the stage of primordial follicles onward; their origin is also unclear with two possible sources, which are the adjacent mesonephros and the ovarian stroma.

GENETIC MECHANISMS

The Importance of the Y Chromosome and the SRY Gene

Compelling evidence for the importance of the Y chromosome in testicular differentiation, irrespective of the number of X chromosomes present, has existed since 1959 when the 47,XXY karyotype for Klinefelter syndrome and the 45,X karyotype for Turner syndrome were first reported. Nonetheless, the testis-determining region on the Y chromosome was not identified until 1990, when *SRY* (Sex-determining region on the Y) was cloned in man and mouse.

SRY belongs to a family of DNA-binding proteins bearing a high mobility group (HMG) box and its gene maps to Yp11.2, very close to the **pseudoautosomal region 1 (PAR1)** (Figure 14–4), one of the only two regions of the Y chromosome that undergo meiotic recombination with homologous sequences (also known as crossing over) of the X chromosome during male spermatogenesis. The proximity of *SRY* to PAR1 makes it susceptible to aberrant translocation to the X chromosome, providing an explanation for XX maleness owing to SRY translocation.

The transient expression of *SRY* during embryogenesis triggers a cascade of gene interactions that leads to testicular development from the undifferentiated and bipotential gonadal ridge (see Figure 14–3). *SRY* expression initiates between days 41 and 44 post-fertilization in humans in the middle of the gonad and expands toward the poles. The timing and level of *SRY* expression are critical for normal testis differentiation; in fact, decreased or delayed expression results in testicular dysgenesis or the formation of ovotestes in mice. The factors regulating *SRY* expression are being elucidated: SF1, SP1, and the +KTS splice variant of WT1 activate *SRY* transcription. The cofactor CITED2 acts with WT1 and SF1 to increase SRY levels to the critical threshold for triggering testis development. Interaction between GATA4 and its cofactor FOG2 are also required for normal SRY expression and testicular differentiation in mice, possibly through its interaction with the +KTS isoform of WT1. Alternatively, GATA4 could act directly on the *SRY* promoter, after a series of phosphorylation reactions involving GADD45G and MAP3K4.

Other Pathways in Testicular versus Ovarian Differentiation

Considerable progress has been made since SRY was identified, and it has become clear that sex determination is a far more complex process than initially thought, regulated by competing molecular pathways in the somatic cells of the bipotential gonad.

SOX family genes   The SRY-box gene *SOX9* is the earliest upregulated gene in the testis pathway downstream of SRY, followed by CITED4 and other members of the SOX family,
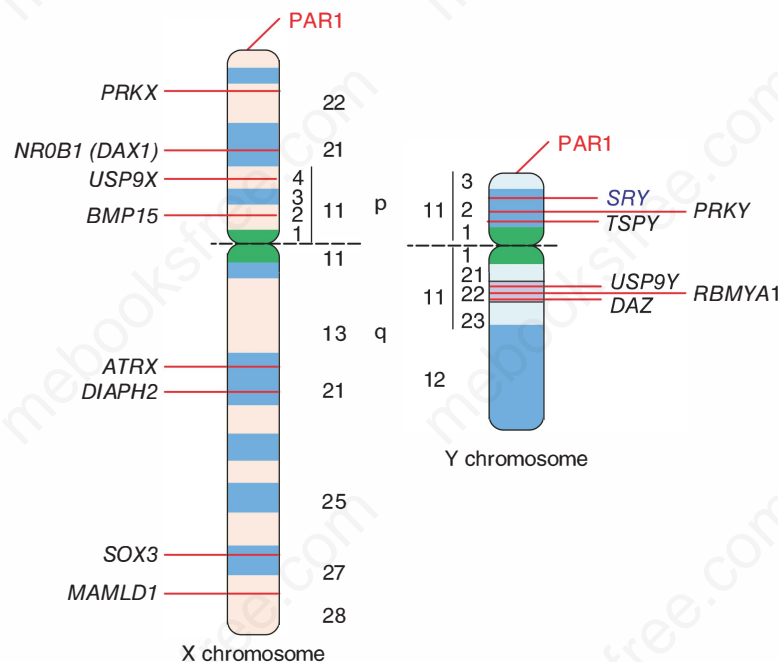
FIGURE 14–4 Sex chromosomes and genes involved in gonadal development and/or function. PAR, pseudoautosomal region.

including *SOX3*, *SOX10*, and *SOX13*, and many other genes that are critical for testicular differentiation. SOX9 is located on chromosome 17q, is an autosomal member of the HMG-box protein superfamily, and is the master regulator of testis differentiation, once SRY has upregulated its expression during a narrow critical time window. SRY and SF1 regulate SOX9 expression by directly binding to several sites within a 3.2-kb testis-specific enhancer (TES) or 1.4-kb of its core element (TESCO), present in its distal promoter. SOX9 is able to maintain its own expression via two different mechanisms: by interacting with SF1 to activate TES and with the help of feed-forward enhancing pathways involving FGF9 and PGD2. These mechanisms succeed in triggering Sertoli cell differentiation. FGF9 and SOX9 also up-regulate AXIN1 and GSK3β, which promote the destabilization of β-catenin and, thus, serve to block ovarian development (see Figure 14–3).

SOX9 mimics SRY independently of SRY expression: for instance, SOX9 overexpression during early embryogenesis induces testicular differentiation in XX humans and mice. SOX9 also affects the differentiation of the reproductive tract by upregulating AMH expression. SOX9 is not testis-specific: it is also expressed in chondrocytes, bile duct, CNS, hair follicles, heart, lung, pancreas and retina, which explains the several malformations observed in patients with SOX9 deficiency. Duplication of SOX9 is a cause of 46,XX sex reversal and SOX9 deletion causes campomelic dysplasia—a severe type of skeletal dysplasia associated with 46,XY sex reversal.

SOX3, encoded by a single-exon, X-linked gene, is involved in brain, pituitary and craniofacial development. Although SOX3 is not required for normal testicular differentiation, its ectopic overexpression in XX gonadal ridges induces testicular differentiation, probably by acting synergistically with SF1 to upregulate SOX9 expression.

SOX8 and **SOX10** are two autosomal SOX family members expressed in the gonads and in several other tissues. During mouse embryo development, the expression of SOX8 and SOX10 is triggered shortly after that of SOX9, but at lower level. Like SOX9 itself, SOX8 and SOX10 can synergize with SF1 and up-regulate SOX9 (see Figure 14–3) upon binding to TESCO. SOX8 can bind the canonical target DNA sequences and activate AMH transcription acting synergistically with SF1, but with less efficiency than SOX9. Later during fetal development, an interaction between SOX9 and SOX8 is required for basal lamina integrity of testis cords and for suppression of FOXL2, two events essential events to the normal development of testis cords.

Other testicular morphogens   The signaling cascade initiated by SRY and SOX family factors counteract the WNT4/FST/β-catenin pathway. When the former prevails, activin B, VEGF, PDGF, and other unidentified factors facilitate endothelial migration and formation of coelomic vessel and the testis-specific vasculature.

The formation of the seminiferous cords, which enclose the germ cells finally giving rise to the gametes, requires the formation of a basal lamina around Sertoli and germ cells. The presence of peritubular cells is essential. The differentiation of these cells

seems to be triggered by Sertoli cell-secreted factors, like DHH. Peritubular cells assemble around the cords and interact with Sertoli cells to produce the proteins of the extracellular matrix, like collagen, fibronectin and laminin, the components of the basal lamina.

First genetic theories about ovary differentiation   Molecular pathways underlying fetal ovarian differentiation remain largely unexplored. Most of the present knowledge relies on observations concerning the absence of testicular differentiation. However, the "default pathway theory" proposing that the sole absence of *SRY* results in ovarian differentiation seems to apply to specific experimental conditions in rodents. In humans, *SRY* mutations or deletions result in gonadal dysgenesis, but not in ovarian differentiation as no oocyte meiotic progression or follicle development has been described in these conditions, even during fetal life. Recent findings indicate that the coordinated action of several factors is needed for the differentiation and stabilization of the ovaries.

WNT/RSPO1/β-catenin and FST signaling pathways   Ovary-specific gene expression is triggered somewhat later than testis-specific genes. Despite similar expression in the XX and XY gonadal ridges, some factors, such as WNT4, R-Spondin 1 (encoded by *RSPO1*), β-catenin (encoded by *CTNNB1*) and follistatin (encoded by *FST*), increase their expression in the XX gonad. RSPO1 upregulates the expression of WNT4 and both factors work synergistically to counteract SOX family proteins, and to stabilize β-catenin, which accumulates in the nucleus and interacts with LEF1 leading to the activation of downstream genes and further inhibition of SOX9. *FST*, one of the genes up-regulated by the WNT4-triggered pathway, antagonizes activin B, thereby blocking the formation of the testis-specific vasculature (see Figure 14–3), an essential early step for testicular differentiation. Apart from this anti-testis action, WNT4 also has a pro-ovarian effect by supporting germ cell survival. The expression of all these factors is not limited to the developing ovary; for instance, WNT4 is expressed in the brain, lung, and mesonephros; RSPO1, in the CNS, lung and skin; and β-catenin, and follistatin are ubiquitous.

FOXL2 and follicular morphogenesis   The forkhead transcription factor FOXL2 is one of the earliest markers of ovarian differentiation and is essential for follicle formation: it can be detected in pregranulosa cells, and female mice lacking *Foxl2* have follicles whose maturation is arrested between the primordial and primary stages, with subsequent atresia. FOXL2 is also involved in the development of the eyelids.

MAP3K1 modulates the balance between female and male pathways   As explained, SOX9 and FGF9 upregulate AXIN1 and GSK3β, which promote the destabilization of β-catenin and block ovarian development. MAP3K1 sequestrates AXIN1, thus promoting the stabilization of β-catenin and favoring the ovarian pathway.

DMRT1 and DAX1: their precise roles remain elusive   The DM domain transcription factor family member DMRT1 appears

to play a conserved role in male gonad development among verte-brates. In mice, DMRT1 is required for normal testicular mor-phogenesis; loss of DMRT1 expression activates FOXL2 and reprograms Sertoli cells into granulosa cells, even in postnatal life, suggesting that DMRT1 is essential to maintain mammalian testis differentiation throughout life. In humans, the role of DMRT1, which maps to 9p24.3, is less clear although it seems to play a role in testicular differentiation, as suggested by observations in patients with 9p deletions. DAX1, encoded by *NR0B1*, is an orphan nuclear receptor mapping to the DSS (Dosage Sensitive Sex-reversal) region on Xp21. DAX1 was the first putative testis repressor and/or ovarian determining gene, as suggested by the observation that a duplication of DSS results in sex-reversal in 46,XY patients, and DAX1 overexpression in transgenic XY mice impairs testis differentiation by antagonizing the ability of SF1 to synergize with SRY action on SOX9. However, a *Dax1* gene knockout did not prevent ovarian differentiation in XX. Both low and high DAX1 expression results in abnormal gonadal differen-tiation. DAX1 is also involved in the development of the adrenal cortex. Losses of function DAX1 mutations are associated with congenital adrenal hypoplasia and hypogonadism in males.

DIFFERENCES IN TESTICULAR AND OVARIAN GERM CELL DEVELOPMENT

As mentioned, primordial germ cells are bipotential. Interest-ingly, their commitment to a spermatogenic or an oogenic fate does not depend on their XY or XX karyotype. Rather, the functional characteristics of the surrounding somatic cells of the gonads regulate germ cell sexual differentiation. In the female embryo, germ cells are exposed to high levels of **retinoic acid** secreted by cells from the mesonephros and the ovary expressing retinaldehyde dehydrogenases (ALDHs, also known as RALDHs). Exposure to retinoic acid induces the expression of **STRA8**, which triggers germ cell meiosis. Thus, oogonia give rise to primary oocytes, which become surrounded by a single layer of flat granulosa cells to form primordial follicles. Granulosa cells become cuboid and proliferate, giving rise to primary and small growing follicles. During fetal life meiosis progresses but is arrested at the diplotene stage of the first mei-otic prophase until puberty.

In male embryos, primordial germ cells colonizing the testis become clustered in the testis cords, surrounded by Sertoli cells; while Leydig cells are present in the interstitial tissue between cords. Germ cells proliferate by mitosis but do not enter meiosis because the cytochrome P450 enzyme CYP26B1 expressed by Sertoli and Leydig cells mediates the degradation of retinoic acid and hampers the expression of STRA8. In Ser-toli cells, CYP26B1 expression is activated by SOX9 and SF1. FGF9, another factor upregulated in Sertoli cells in response to SOX9, also prevents germ cells from entering meiosis. Thus, germ cells give rise to gonocytes that differentiate to spermato-gonia and proliferate by mitosis, but meiosis only starts at puberty in the male gonad.

HORMONE-DEPENDENT DIFFERENTIATION OF THE GENITALIA

ONE GONAD, TWO CELLS, TWO HORMONES

As already mentioned, Jost's experiments on fetal endocrinology provide the basis for the comprehension of sexual differentiation: only the testis—not the ovary—plays a role in sexual differentia-tion of the internal and external genitalia during fetal life. He went further to show that the testicular hormone secretion had a dual nature, since androgens alone were not capable of fulfilling all features of sex differentiation. We now know that androgens, secreted by Leydig cells, are responsible for the development of the unipotential Wolffian ducts into the male internal reproductive ducts and also for the virilization of the bipotential urogenital sinus and external genitalia, and that AMH, secreted by Sertoli cells, drives the regression of the unipotential Müllerian ducts (see Figure 14–3). In the absence of testes and, therefore, of androgen and AMH action, Wolffian ducts regress, Müllerian ducts develop into the internal female ducts and external genitalia follow the female pathway. In Jost's experiments with early castration of rab-bit embryos, all fetuses developed female internal and external genital structures indicating that ovarian hormones play no role in normal fetal sex differentiation. Other experimental models and human disorders have confirmed this observation.

AMH AND THE FATE OF MÜLLERIAN DUCTS

Anti-Müllerian hormone (AMH), a member of the TGF-β fam-ily, is responsible for the regression of Müllerian ducts in the male fetus. AMH is expressed at high levels by Sertoli cells as soon as they differentiate. Although Müllerian duct regression takes place in the eighth and ninth weeks of fetal life, Sertoli cells continue to produce high amounts of AMH until puberty. In the female, AMH is produced at lower levels by granulosa cells of growing follicles from the twenty-third fetal week, a stage where Müllerian ducts are no longer responsive to it.

Regulation of AMH Expression

AMH transcription in fetal Sertoli cells is triggered by SOX9. SF1, GATA4, and WT1 synergize to increase SOX9-activated AMH transcription, a synergy antagonized by DAX1. This initial stage of AMH production is independent of gonadotropins. Subse-quently, during the second and third trimesters of fetal life and in the neonate, FSH increases testicular AMH production by induc-ing Sertoli cell proliferation and by upregulating AMH transcrip-tion in each Sertoli cell. AMH declines at puberty, when Sertoli cells mature; this is an androgen-mediated effect. Remarkably, AMH expression is not inhibited in fetal and neonatal periods in

spite of high androgen production by the testes; this is due to the fact that Sertoli cells do not express the androgen receptor in those periods of life.

AMH Action

AMH signals through two distinct membrane receptors present in the mesenchymal cells that surround the Müllerian duct. Following proteolytic cleavage into its N and C terminal fragments, AMH binds specifically to the AMH receptor type 2, which subsequently activates a non-specific type 1 receptor. AMH action on Müllerian ducts provokes apoptosis in some epithelial cells whereas others undertake epithelial-to-mesenchymal transformation. Consequently, the epithelial Müllerian ducts disappear beginning days 55 to 60 of fetal life. By the end of the ninth week, Müllerian ducts are no longer responsive to AMH; therefore, any delay in AMH expression may result in the persistence of Müllerian derivatives.

Müllerian Derivatives in the Female

In the absence of AMH action, Müllerian ducts persist and develop into an epithelial tube that gives rise to the endometrium, while the surrounding mesenchyme differentiates into the myometrium of the uterus and Fallopian tubes. Müllerian derivatives also contribute to the upper portion of the vagina (see Figure 14–3).

ANDROGENS AND THE FATE OF THE WOLFFIAN DUCTS, UROGENITAL SINUS, AND EXTERNAL GENITALIA

The interstitial compartment of the testis contains developing Leydig cells, the most important androgen-producing cells in the male, from the beginning of the ninth week in the human embryo. The origin of Leydig cells is still a matter of debate: the precursors have been proposed to be either migrating cells from the coelomic epithelium, the mesonephros or the neural crest, or resident cells present in the adreno-gonadal primordium. Leydig cells also produce insulin-like factor 3 (INSL3), involved in the transabdominal phase of testicular descent. The initial differentiation of fetal Leydig cells depends on Sertoli cell-secreted PDGFs, but further Leydig cell differentiation, proliferation, and steroidogenic capacity depend on placental hCG in the first and second trimesters of gestational life, and on fetal pituitary LH thereafter. Both hCG and LH act on the same LH/CG receptor, a G protein-coupled receptor present in testicular Leydig cells. Testosterone levels peak between 11 and 14 weeks in the male fetus.

Steroidogenesis

Testosterone production
Testosterone is the most abundant androgen produced by the testis. Steroidogenesis is a long process starting from cholesterol through the coordinated action of steroidogenic enzymes (Figure 14–5), most of which are also expressed in the adrenal gland, explaining why many steroidogenic disorders are common to the testis and adrenal. The initial step in steroidogenesis is the conversion of cholesterol to pregnenolone, mediated by the P450 side-chain cleavage enzyme (P450scc) in the inner mitochondrial membrane. The transfer of cholesterol into the mitochondria is dependent on steroidogenic acute regulatory protein (StAR) and TSPO. Pregnenolone is subsequently metabolized by P450c17 that has two distinct activities: a 17α-hydroxylase activity responsible for the conversion of pregnenolone to 17α-hydroxypregnenolone and a 17,20 lyase activity, capable of converting 17α-hydroxypregnenolone to dehydroepiandrosterone (DHEA). The flavoprotein P450 oxidoreductase (POR) and the cytochrome b5 are required for optimal 17,20 lyase activity. P450c17 and its partner proteins also convert the Δ4 compound progesterone into 17α-hydroxyprogesterone and Δ4-androstenedione. The final testicular enzyme in testosterone biosynthesis is 17β-hydroxysteroid dehydrogenase type 3 (17β-HSD-3), which converts Δ4-androstenedione to testosterone and the Δ5 steroid DHEA to androstenediol. Several 17β-HSD isoforms have been identified, but only the type 3 is expressed in the testis. Another hydroxysteroid dehydrogenase, 3β-HSD type 2, is responsible for the conversion of Δ5 to Δ4 steroids. All these enzymes, except 17β-HSD-3, are present in both the testis and the adrenal cortex.

Dihydrotestosterone (DHT) production
Testosterone is less potent than its metabolite DHT, which binds the androgen receptor with higher affinity than testosterone, and also cannot be aromatized to estrogen, thus its effects are purely androgenic. In the classic pathway, testosterone is converted to DHT in target tissues by the enzyme 5α-reductase type 2 expressed in the urogenital sinus, including the prostate primordium, and in the fetal genital skin. Another isoenzyme, 5α-reductase type 1, is expressed in the adrenal cortex, in liver after birth and in skin from the time of puberty. An alternative, also called "backdoor," pathway exists for DHT production which does not involve testosterone as a precursor (see Figure 14–5). Progesterone and 17-hydroxyprogesterone are respectively converted to dihydroprogesterone and 17-hydroxydihydroprogesterone by 5α-reductase type 1, and to allopregnanolone and 17-hydroxyallopregnanolone by 3αHSD. P450c17 convert them to androsterone. The latter is subsequently finally metabolized to DHT via androstanediol or androstenedione by 17β-HSD types 3 or 5, and 17β-HSD type 6 together with 3αHSD.

Estrogen production
Estrogens are synthesized from androgens by the action of the P450 aromatase, which is expressed in the granulosa cells of the ovary, the placenta and also in muscle, liver, hair follicles, adipose tissue, and brain.

Adrenal cortex steroids
The adrenal cortex comprises the zona glomerulosa, which synthesizes mineralocorticoids (aldosterone) essentially under the regulation of the renin/angiotensin system and serum potassium concentration, the zona fasciculata, responsible for glucocorticoid (cortisol) secretion primarily under the stimulation of ACTH, and the zona reticularis, where adrenal androgens (DHEA sulphate and Δ4-androstenedione) are

FIGURE 14–5 This diagram illustrates adrenal, ovarian, and testicular steroidogenic pathways. Under the usual circumstances in humans, 17-OHP is a poor substrate for CYP17A1 (dotted line). In CAH due to 21-OHD, the defect leads to accumulation of progesterone and 17-OHP, which are diverted not only to AD and T but also to additional metabolites such as 16α-hydroxyprogesterone and 21-deoxycortisol. The adrenal cortex also secretes 11β-hydroxyandrostenedione (11OHAD), and 11-ketotestosterone (11KT). In green, steps that are common to adrenals and gonads; in red, those related to gonadal steroids, in dark gray, those related to adrenal steroids, and in blue those involved in the "backdoor pathway." Δ5 pathway involving pregnenolone through DHEA, and Δ4 pathway involving progesterone through testosterone are indicated. Δ5 steroids are converted to Δ4 steroids by 3β-HSD. 3β-HSD, 3β-hydroxysteroid dehydrogenase; 5α-red, 5α-reductase; 17-HSD, 17β-hydroxysteroid dehydrogenase; P450aldo, Cytochrome P450 aldosterone synthase; P450aro, Cytochrome P450 aromatase; P450c11, Cytochrome P450 11β-hydroxylase; P450c17, Cytochrome P450 17α-hydroxylase/17,20-lyase; P450c21, Cytochrome P450 21-hydroxylase; P450scc, Cytochrome P450 steroid sidechain cleavage; AKR, Aldo-keto reductase; DHEA, Dehydroepiandrosterone; DHT, Dihydrotestosterone; DOC, Desoxycorticosterone; POR, Cytochrome P450 oxidoreductase; SF1, Steroidogenic factor 1; StAR, Steroidogenic acute regulatory protein. (Reproduced with permission from Rey RA, Grinspon RP. Normal male sexual differentiation and aetiology of disorders of sex development. *Best Pract Res Clin Endocrinol Metab*. 2011 Apr;25(2):221-238.)

produced under ACTH control. The initial steps are shared with gonadal steroidogenesis (see Figure 14–5). The existence of the cytochrome P450 with 21-hydroxylase activity converts progesterone to deoxycorticosterone in the zona glomerulosa and 17-hydroxyprogesterone to deoxycortisol in the zona fasciculata. Subsequently, another cytochrome P450, aldosterone synthase, converts deoxycorticosterone to corticosterone and aldosterone in two successive steps, whereas 11β-hydroxylase converts deoxycortisol to cortisol. In the fetus, the zona reticularis makes little Δ4-androstenedione and massive amounts of DHEAS, because expression of 3β-HSD type 2 is low as compared to that of DHEA-sulfotransferase. With adrenarche, the zona reticularis secretes increased amounts of DHEAS, DHEA, and androstenedione. Testosterone production is also low because 17β-HSD type 3 is not expressed. Alternative enzymes of the AKR1C family, including 17β-HSD type 5 (encoded by *AKR1C3*) are probably responsible for the small amounts of testosterone production.

Androgen Action in Target Tissues

All androgens act through binding with the same androgen receptor, a zinc-finger transcription factor encoded by the *AR* gene mapping to Xq12. Like other nuclear receptors, the androgen receptor is composed of three major functional domains: a large N-terminal domain that precedes the DNA-binding domain, which is connected to the C-terminal ligand-binding domain by a short-hinge region. Unlike other members of the steroid receptor family, the androgen receptor resides mainly in the cytoplasm and translocates into the nucleus in the presence of its ligand. The androgen receptor binds to specific DNA motifs, the androgen response elements (ARE), present in the promoter regions of target genes. Interestingly, the *SRD5A2* gene encoding 5α-reductase type 2 has ARE sequences in its promoter and is regulated by testosterone. The androgen receptor activity is modulated by coregulators, partner

Wolffian Duct Derivatives

The Wolffian ducts are exposed to very high local concentrations of testosterone, unlike more distant organs receiving testosterone only via the peripheral circulation. Therefore, conversion to DHT does not seem essential for their stabilization and differentiation (see Figure 14–3). Between weeks 9 and 13, the upper part of the Wolffian duct differentiates into the epididymis. Below, it becomes the vas deferens, which opens into the urogenital sinus at the level of Müllerian tubercle. The seminal vesicle originates from a dilatation of the terminal portion of the vas deferens in 12-week-old fetuses. In the absence of androgen action, for example, in the normal XX fetus or in cases of androgen synthesis or androgen receptor defects, the Wolffian ducts degenerate by weeks 10 to 12.

The Bipotential Urogenital Sinus

The urogenital sinus originates the bladder of both sexes, the prostate and the prostatic and membranous portions of the urethra in the male, and the whole urethra and the lower portion of the vagina in the female. Male differentiation is dependent on testosterone conversion to DHT.

Prostatic buds appear at approximately 10 weeks at the site of the Müllerian tubercle and grow into solid branching cords. Maturation of the prostatic gland is accompanied by development of the prostatic utricle. Two buds of epithelial cells, called the sinoutricular bulbs in the male, develop from the urogenital sinus close to the opening of the Wolffian ducts and grow inward, fusing with the medial Müllerian tubercle, to form the sinoutricular cord, enclosed within the prostate gland, which canalizes at 18 weeks to form the prostatic utricle, the male equivalent of the vagina.

Female differentiation of the urogenital sinus is characterized by the acquisition of a separate vaginal opening on the surface of the perineum (Figure 14–6). At the end of the ambisexual stage, the vaginal anlage is located just underneath the bladder neck. In females, the lower end of the vagina slides down along the urethra until the vaginal rudiment opens directly on the surface of the perineum at 22 weeks. The hymen marks the separation between the vagina and the diminutive urogenital sinus, which becomes the vestibule.
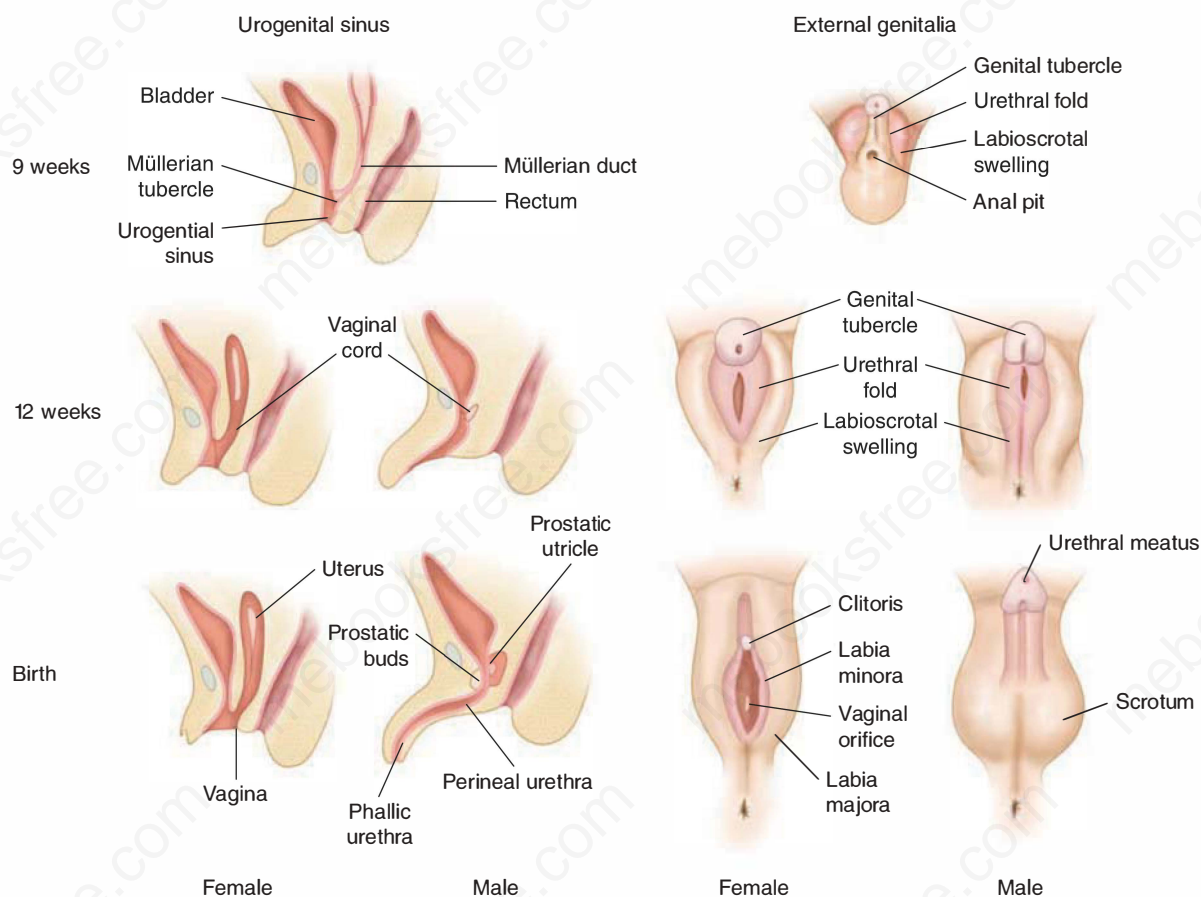
FIGURE 14–6 Sex differentiation of urogenital sinus (left) and external genitalia (right). See text for details. (Reproduced with permission from Rey R, Josso N. Sexual differentiation. In: De Groot LJ, Beck-Peccoz P, Chrousos G, et al. Endotext [WWW.ENDOTEXT.ORG]. Published by MDTEXT.COM,INC, 2016. South DARTMOUTH,MA 02748.)

The embryology of the vagina is still matter of debate. The most accepted view proposes that the upper part of the vagina derives from the Müllerian ducts and the lower part from the sinovaginal bulbs, which by fusion form the vaginal plate, derived from the urogenital sinus. The Wolffian ducts do not seem to contribute cells to the sinovaginal bulbs but they may have a helper function during downward movement of the vaginal bud in the female.

The Bipotential External Genitalia

Masculinization of the external genitalia begins during the ninth fetal week in response to DHT, and is characterized by lengthening of the anogenital distance. Subsequently labioscrotal folds fuse in a dorsal to ventral fashion (see Figure 14–6). The embryological development of the penile urethra is controversial. Some authors describe it as a two-step process, with the proximal urethra forming by fusion of the urethral folds around the urethral plate and the distal urethra arising from an invagination of the apical ectoderm. Others propose that the entire male urethra is of endodermal origin, formed by the urethral plate dorsally and the fused urethral folds ventrally.

The genital tubercle gives rise to the glans penis and the corpora cavernosa. Surprisingly, there is no size difference between penile or clitoral size until 14 weeks despite the fact that serum testosterone levels peak between 11 and 14 weeks in males. Urethral organogenesis is complete at 14 weeks, apart from a physiological ventral curvature, which can persist up to 6 months of gestation. Maximal fetal penile growth occurs during the third trimester, at a time when male testosterone levels, now under fetal LH control, are declining.

In the female fetus, the anogenital distance does not increase; the urethral folds do not fuse and form the labia minora, while the labioscrotal swellings give rise to the labia majora. The dorsal commissure forms at their junction. When the vagina acquires a separate perineal opening, the diminutive pars pelvina and the pars phallica of the urogenital sinus become the vestibule. The genital tubercle emerges as a ventral and medial swelling and becomes the origin of the clitoris.

Testicular Descent

During fetal life, the testis migrates from its initial pararenal location to its final position in the scrotum. Testicular descent can be divided into two phases. Initially, the upper pole of the testis is connected to the posterior abdominal wall by the cranial suspensory ligament while a primitive gubernaculum extends from the caudal pole to the inner inguinal ring. During the transabdominal phase, between fetal weeks 10 and 15, the cranial suspensory ligament dissolves and the gubernaculum testis swells and pulls the testis down to the inguinal ring. This phase is dependent on androgens and INSL3. The inguinoscrotal phase occurs between 27 and 35 weeks, and is mainly dependent on androgen action. The clinical relevance of the involvement of other factors, like homeobox A10 (HOXA10), calcitonin gene-related peptide (CGPR), and AMH, still needs to be fully elucidated. Physiological cryptorchidism is frequent in premature infants. In the female, the cranial ligament holds the ovary in a high position, and the gubernaculum, giving rise to the round ligament, remains long and thin.

DISORDERS OF SEX DIFFERENTIATION (DSD)

DEFINITIONS AND HISTORICAL PERSPECTIVES

Terms that have been used to describe these conditions include the following:

1. **Ambiguous genitalia**—the condition in which an infant's external genitals do not appear clearly male or female.
2. **Pseudo-hermaphroditism**—the condition in which external genitalia are not consistent with gonadal development (ovaries or testes) and karyotype. These were formerly classified based on karyotype in which the pronoun used (female or male) described the karyotype.
   a. Female pseudo-hermaphroditism—indicating masculinized external genitalia in an individual with ovaries, 46,XX karyotype and female internal genitalia.
   b. Male pseudo-hermaphroditism—indicating incompletely virilized genitalia in an individual with testes and 46,XY karyotype.
3. **True hermaphroditism**—an individual with both ovarian and testicular tissue.
4. **Intersex**—a condition of incomplete or disordered genital or gonadal development leading to a discordance between genetic sex, gonadal sex, and genital sex.
   a. This definition preceded the terminology: DSD, first proposed in 2006.
   b. Examples vary from ambiguous genitalia, to male genitalia with 46,XX karyotype or female genitalia with 46,XY karyotype, to male genitalia and internal ovaries or female genitalia and testes.
5. **Disorders of sex development**—these are congenital conditions in which development of chromosomal, gonadal and genital sex is atypical.
   a. This newly created broad category was suggested to replace the term intersex.
   b. Categories included the following plus other rare classifications (Table 14–1):
      i. 46,XY DSD replaced the male pseudo-hermaphrodite category and includes undervirilized and under-masculinized XY male.
      ii. 46,XX DSD replaced the female pseudo-hermaphrodite category: masculinized and over-virilized 46,XX female.
      iii. Sex chromosome DSD as an inclusive category including the former true hermaphrodite and mixed gonadal dysgenesis or asymmetric gonadal differentiation categories (45,X/46,XY), 46,XX/46,XY chimeric, ovotesticular DSD and also Turner syndrome and Klinefelter syndrome and its variants.
      iv. To accommodate the rare conditions such as XY female, XX sex reversal, and XY sex reversal, the terms 46,XX testicular DSD, and 46,XY complete gonadal dysgenesis were proposed.
   c. This broad DSD definition has been interpreted to include conditions such as cryptorchidism and isolated hypospadias, sex chromosome aneuploidies such as Turner and

TABLE 14–1 Karyotype-based classification of DSD.

DSD Type	Pathogenic Classification	Etiology
46,XY DSD	Malformative DSD	Defective morphogenesis of the genital primordia
	Dysgenetic DSD	Complete, partial, or asymmetric gonadal dysgenesis
		Ovotesticular DSD
	Non-dysgenetic DSD	Disorders of androgens synthesis
		Disorders of androgen action (AIS)
		Disorders of AMH synthesis or action
46,XX DSD	Malformative DSD	Defective morphogenesis of the genital primordia
	Dysgenetic DSD	Ovotesticular DSD
		Testicular DSD
		Ovarian dysgenesis
	Non-dysgenetic DSD	Excessive adrenal androgen synthesis (CAH)
		Decreased placental androgen aromatization (aromatase deficiency)
		Androgen-secreting tumors
		Androgenic drugs
Sex chromosome DSD	Dysgenetic DSD	Ovotesticular DSD (46,XX/46,XY, mosaicisms with 46,XX and 46,XY lineages)
		Asymmetric gonadal dysgenesis (45,X/46,XY, other mosaicisms with a 45,X lineage)
		Turner syndrome (45,X and variants)
		Klinefelter syndrome (47,XXY and variants)
		Triple X syndrome (47,XXX and variants)

Klinefelter syndromes, as well as adrenal enzyme deficiencies to include 46,XY individual having appropriate internal and external genital differentiation, such as congenital adrenal hyperplasia (CAH), 21-hydroxylase deficiency.

  i. This new terminology has been criticized for carrying a stigma and for unnecessarily medicalizing those with intersex traits.

  ii. The breadth of the new terminology requires further definition to clarify what category is being discussed and also obscures the incidence of subcategories, for example, the incidence of DSD individuals with external genitalia so ambiguous that it is necessary to determine sex of rearing.

6. **Sex and gender** (definitions may differ or overlap)

  a. Sex—defined traditionally as either male or female based on their reproductive functions. To differentiate from gender, sex refers to biology including appearance of external genitalia, traditionally considered a binary classification.

  b. Gender—the behavioral, cultural, or psychological traits typically associated with male or female sex; classified as feminine, masculine or neuter.

  c. Gender identity—an individual's self-designation as male or female (or uncommonly a non-binary definition). Gender identity develops over time.

  d. Gender role—sex-typical behaviors are impacted by hormonal exposure, particularly androgens and additionally defined by societies.

  e. Sexual orientation—the erotic interest as heterosexual, homosexual, bisexual, or fluid.

  f. Sexual identity—may be considered a composite of gender identity, gender role, and sexual orientation.

Since ancient times, society has been challenged with how to respond to individuals with anatomic differences of genital development. The anatomic variations influenced the characters of ancient mythology, that is, hermaphrodites. Responses to genital atypia continue to evolve, involving changes in perceptions, recommendations, and guidelines for male or female assignment for the infant born with ambiguous genitalia.

Historically, the approach to the individual with ambiguous genitalia has been based upon then current understanding of both the pathophysiology, the determinants of gender and the cultural background of the time. In spite of the advancement, major gaps in knowledge persist, particularly regarding the determinants, hence predictors, of gender development. When considering progression of the medical approach to sex assignment as "generations," the first generation, used prior to the mid-20th century, was based on the perception that "true sex" was ascertained by gonadal differentiation and that "true gender" would follow "true sex." Hence, gonadal histology became the basis for sex assignment or reassignment. Thus, if ovarian development was documented, sex assignment was female, regardless of the extent of

masculinization of the reproductive system, and the converse was true for those with testicular histology, who received a male assignment. This paradigm was modified during the 1950s after chromosomal sex could be determined. The perception of sex chromosomes, before appreciation of the role of genes on both autosomes and sex chromosomes on sex differentiation, was that a Y chromosome meant a male sex, while two Xs meant female. Chromosomal and gonadal sex together determined "true sex" for assignment or reassignment. Tragic outcomes of the practice of recommending a female reassignment for a child with ovaries who essentially fully developed male genitalia and was being raised as male illustrate the fallacy of this approach.

In the 1960s, the next or "second generation" perspective developed in response to the realization that gender development was a complex, multifactorial, although incompletely understood, process. This "Optimal Gender" approach, recognized gender development as a multifactorial process involving both pre- and postnatal variables, heavily influenced by psychosocial factors. Both heredity and environment were felt to be involved. This concept led to the hypothesis that gender was malleable at birth, was significantly influenced by postnatal inputs, and could be imposed by aligning social inputs with sex assignment. Assignment was based upon perceived potential for fertility and capacity for penetrative intercourse, which, together with external genitalia consistent with the sex of rearing, were considered the key determinants for a good adult Quality of Life (QoL). For 46,XX patients, primarily those with 21-hydroxylase deficiency CAH, a female assignment was based on the conviction that suppression of adrenal androgen excess would allow an excellent potential for fertility, and that feminizing genital surgery would provide the capacity for traditional intercourse, and together these factors would form the basis for success in adult life. The availability of cortisone therapy to suppress excessive adrenal androgen secretion supported the assumption that such therapy would be followed by complete feminine development and fertility.

For 46,XY patients with incomplete masculinization, the optimal gender approach criteria included the expectation of poor potential for both fertility and testosterone secretion, so expected adequacy of penis development for penetrative intercourse in adulthood was considered the primary key factor for sex assignment. When an anatomically correct but very small penis was felt to have insufficient growth potential or when surgery for phallus with hypospadias could not be expected to result in an adequately sized and configured penis, a female assignment was considered. The basis of this perspective was apparently the crippling psychological outcome of a few adult males with a very diminutive penis size and the perspective that it would be inappropriate to raise a child as a male who would be incapable of traditional sexual intercourse.

Sex assignment for patients with true hermaphroditism and mixed gonadal dysgenesis (or asymmetric gonadal differentiation) was based upon external and internal reproductive system development and gonadal differentiation. For both diagnoses, if a single testosterone-producing testis was present alongside a phallic structure that could be repaired into a reasonable penis, male assignment was recommended even though fertility was not expected.

Since MGD does not involve functional ovarian tissue, even though fertility was highly unlikely, male assignment was usual because of masculinization. For ovotesticular DSD, if an ovary was present with follicles, a Fallopian tube on that same side and a uterus differentiated to potentially accommodate a fetus, a female assignment was considered. For both diagnoses, gonadectomy was considered because of risk of malignancy.

Influences that impacted this second generation approach included case reports and patient outcome series, including the dissatisfaction and non-acceptance of sex assignment when sex reassignment was attempted following in response to a missed diagnosis, particularly among those who had been exposed to significant androgen during fetal and post-natal life. There was little appreciation of the trauma inflicted when an individual was told that they were actually the opposite sex from what they presumed. The practice of incomplete disclosure of medical information, also a tenet of the second generation approach, was intended to shelter patients from information that was felt to impede adjustment with the sex assigned but in hindsight was clearly unethical, risked the trauma of late disclosure, and underestimated the resilience of human nature.

Although practice shifts had already begun, the "third generation" version of sex assignment guidelines was defined at the 2005 international consensus conference. Recommendations for sex assignment, based upon the data available at that time, were consistent with the principle that each generation applies current scientific knowledge, using the cultural understanding of the time. A female assignment was considered to be appropriate for all 46,XX CAH patients based upon the evidence that 95% of such patients develop a female gender identity, although the lack of adequate data regarding that subgroup of patients with markedly virilized genitalia (Prader stage 4 and 5) was apparent. Female assignment was also recommended for 46,XY patients having CAIS and an LH receptor defect. Male assignment was recommended for patients with 5α-reductase deficiency and 17-hydroxysteroid dehydrogenase deficiency based upon outcome data indicating that the majority of patients identify as males in adult life. Those in the new Gonadal DSD category diagnosed with ovotesticular DSD and MGD, the recommendation remained as previously, assignment being based upon the composite anatomy judged to be more consistent with the better QoL.

A shift toward a male sex assignment had already begun before 2005 for many 46,XY DSD patients, particularly those without a defined etiologic diagnosis. This was influenced by the outcome among 46,XY individuals with cloacal exstrophy, a model of genital ambiguity in which fetal testicular testosterone secretion is normal. The finding that an impressive portion of the 46,XY infants with cloacal extrophy who were reassigned female and orchiectomized during the early neonatal period later expressed a male gender identity suggested an indelible impact of fetal testosterone on gender development outside of any postnatal testosterone exposure.

It is apparent that the perception that karyotype, rather than specific genes on both sex chromosomes and autosomes, persists and may be detrimental to best judgment regarding sex assignment. Also, as with gender dysphoria among those with normal

biologic development, a broader gender model that goes beyond a simple binary system is more likely to be considered.

There was also a shift in surgery over the last half of the previous century for females with clitoromegaly, likely related to increased appreciation of the female sexual responsiveness. Gradually surgery became reserved for those with more extreme degrees of enlargement using refined techniques that preserve innervation. Also, during the 1990s, there was dissatisfaction by former patients with both non-disclosure of their medical history expressed and genital surgery that had been performed when they were young. While it is unclear what portion of former patients were dissatisfied with their treatment, especially surgery during infancy and childhood, these complaints demanded attention of those caring for these patients. This, plus the recognition of the need to reassess care of DSD patients was the motivation for the meeting of the Consensus Group. These caretakers were confronted with accusations, largely focused against Dr. John Money—the researcher who developed the optimal gender strategy in response to a non-intersex case—regarding the negative impact his approach had on their gender development. When no one else provided counseling to the family of a patient who had a traumatic penile amputation during circumcision (not an intersex condition and not presenting in the neonatal period), Dr. Money suggested a female assignment consistent with the perspective at that time that sex of rearing was a major determinant for gender development, if reassignment occurred before 18 months of age. Ultimately, it became clear that this child developed a **male** gender identity, in contrast to another patient with a similar history of accidental penis amputation who developed a female gender identity. This highly publicized case, together with concerns about female assignment for other 46,XY DSD patients, was likely a major factor in the shift away from female assignment.

There is a contingent of concerned individuals who hold the perspective that genital surgery during infancy is simply cosmetic surgery unless some physically life-threatening condition exists. This position does not take into account the unknown role that external genital anatomy plays in gender identity development. Allowing a child to grow up with genitalia that do not fit sex assignment has been considered as social experimentation in some settings (http://www.isna.org/colombia, accessed 2 Nov 2015). The position that surgery should wait until the patient is old enough to decide could be considered to preclude the parents' rights and responsibilities. Further, there is no good guidance to determine when a person has established a permanent gender or when one is old or mature enough to make a decision of this magnitude. There are examples of those who have decided and had surgery and later regretted it and examples of those who asked for surgery but later decided they no longer wanted surgery.

Currently, with the understanding of how complex gender development is, it is realized that it remains unpredictable in many settings. In the past, the approach to sex assignment may have been overly simplistic and too dogmatic. The challenge for the current approach is to consider the multiple facets involved and to arrive at an individualized care plan that is most likely to succeed. Male or female assignment should be approached from the perspective that it should be consistent with the most probable gender identity development and need not align with a potential surgical outcome or the capacity for traditional sexuality. An overarching goal of DSD management is to promote sex assignment that leads to positive self-esteem and sense of fulfillment.

PATHOGENIC CLASSIFICATION

The characteristics of the sex chromosomes, the gonads and/or the genitalia may be atypical or inconsistent for various reasons. In this section, the pathogenic mechanisms underlying the different conditions included under the definition of DSD will be discussed. For this, we shall refer to the embryologic and fetal processes described in the section "Normal fetal sex differentiation" that did not follow a typical development (Table 14–2).

Malformative DSD: Defects in the Morphogenesis of the Urogenital Primordia

Malformative, non-endocrine-related DSD, are due to defects in the early morphogenesis of the anlagen of the reproductive tract in individuals with normal gonadal differentiation. These disorders are suspected when the genital abnormalities are inconsistent with defective sex hormone secretion or action. For instance, isolated aphallia in a patient with a normally developed scrotum and descended testes is inconsistent with a defect in androgen secretion or action. The normal embryologic process underlying the formation of the genital primordia in early embryogenesis is impaired due to genetic or environmental causes. In these patients, serum levels of gonadal hormones are usually within the normal range.

A. Defective morphogenesis of the Müllerian or Wolffian ducts   Malformations of the vagina and uterus occur in females with hand-foot-genital syndrome due to *HOXA13* mutations and in patients with Mayer-Rokitansky-Küster-Hauser syndrome (MRKH, OMIM 277000), a heterogeneous disorder with abnormalities that may range from upper vaginal atresia to total Müllerian agenesis associated with urinary tract abnormalities, and even cervicothoracic somite dysplasia (known as MURCS association, OMIM 601076). Mutations in the *WNT4* gene have been identified in a small subset of MRKH patients, but do not seem to be the main etiology of MRKH syndrome. Vaginal atresia has been described in McKusick-Kaufman syndrome (OMIM 236700), probably caused by mutations in the *MKKS* gene.

Congenital bilateral absence of the vas deferens is found in 95% of males with cystic fibrosis (OMIM 277180), and accounts for 1% to 2% of cases of male infertility.

B. Defective morphogenesis of the urogenital sinus and the *primordia* of the external genital   Early morphogenetic defects of the cloaca or the genital primordia, occurring independently of hormone action, may be responsible for isolated aphallia or isolated hypospadias in patients with normal gonadal function (Figure 14–7).

TABLE 14–2 Pathogenic classification of disorders of sex development (DSD).

DSD Type	Pathogenic Mechanism	Examples	Implicated Gene	OMIM
Malformative DSD	Defective morphogenesis of the Müllerian ducts or Wolffian ducts	Hand-foot-genital syndrome	HOXA13	140000
		MRKH syndrome	WNT4	277000
		MURCS association		601076
		McKusick-Kaufman syndrome	MKKS	236700
		Absence of the vas deferens	CFTR	277180
	Defective morphogenesis of the urogenital sinus and the primordia of the external genitalia	Isolated aphallia	ROR2	268310
		Isolated hypospadias	GLI3	146510
			SPECC1L	145410
			PITX2	180500
Dysgenetic DSD	Complete or partial gonadal dysgenesis (46,XY)	Swyer syndrome/Partial testicular dysgenesis	SRY	480000
			MAP3K1	613762
			DSS	300018
		Campomelic dysplasia	SOX9	608160
		Testicular dysgenesis ± adrenal insufficiency (SF1 deficiency)	NR5A1	612965
		Denys-Drash syndrome Frasier syndrome	WT1	607102
		Testicular dysgenesis + mental retardation	DMRT1	602424
		ATRX syndrome	ATRX	301040
		Testicular dysgenesis with minifascicular neuropathy	DHH	607080
	Asymmetric gonadal differentiation (Mixed gonadal dysgenesis)	45,X/46,XY		
		Other mosaicisms with 45,X and Y lineage		
		46,XY (rare)		
	Ovotesticular	46,XX/46,XY		
		Mosaicisms with XX and XY		
		46,XY (extremely rare)		
	Ovotesticular/Testicular	46,XX, SRY+	SRY	400045
		46,XX, SRY-	SOX9	608160
			SOX10	602229
			SOX3	313430
			RSPO1	610644
			WNT4	611812
Non-dysgenetic DSD with testes	Disorders of androgen synthesis	Leydig cell aplasia/hypoplasia	LHCGR	238320
		Smith-Lemli-Opitz syndrome	DHCR7	270400
		Lipoid congenital adrenal hyperplasia	STAR	600617
			CYP11A1	118485
		P450c17 deficiency	CYP17A1	202110
		P450 oxidoreductase deficiency	POR	201750
		3β-HSD type 2 deficiency	HSD3B2	201810
		17β-HSD type 3 deficiency	HSD17B3	264300
		DHT synthesis defects	SRD5A2	264600
			AKR1C2	600450
			AKR1C4	600451
		MAMLD1 defects	MAMLD1	300120
	Disorders of androgen action	Androgen insensitivity syndrome	AR	300068

(continued)

TABLE 14–2 **Pathogenic classification of disorders of sex development (DSD). (Continued)**

DSD Type	Pathogenic Mechanism	Examples	Implicated Gene	OMIM
	Disorders of AMH synthesis or action	Persistent Müllerian duct syndrome	*AMH*	600957
			AMHR2	600956
Non-dysgenetic DSD with ovaries	Excessive adrenal androgen synthesis	Congenital adrenal hyperplasia	*CYP21A2*	201910
			CYP11B1	202010
			HSD3B2	201810
			POR	201750
		Glucocorticoid resistance	*NR3C1*	615962
	Impaired placental androgen aromatization	Placental aromatase deficiency	*CYP19A1*	613456
	Androgen-producing tumors	Luteoma, thecoma, etc		
	Androgen/progestogen consumption	Norethindrone, ethisterone, norethynodrel, medroxyprogesterone, danazol		
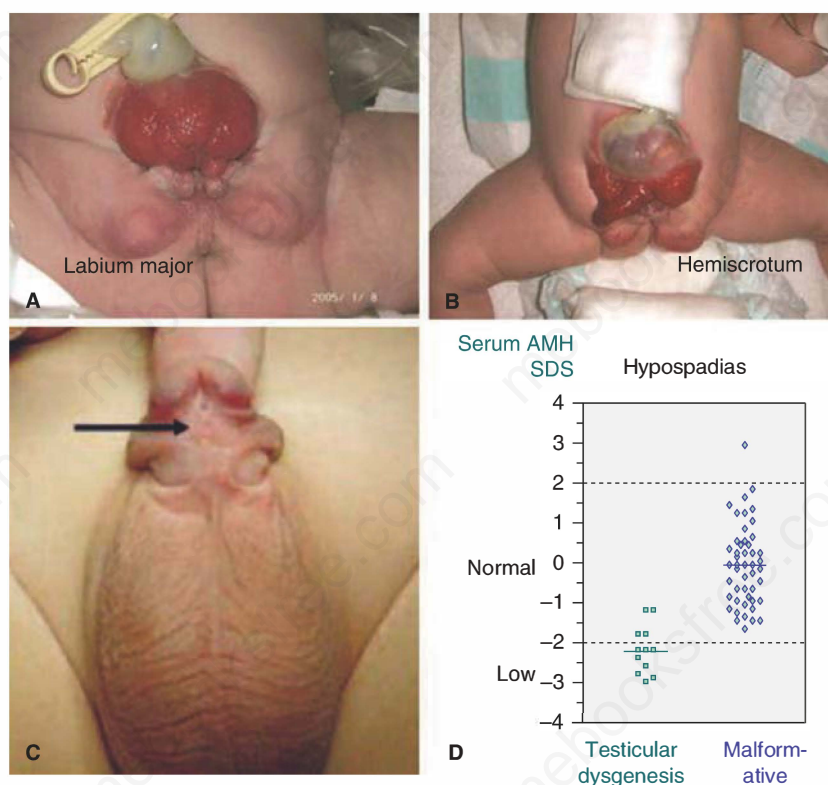
FIGURE 14–7 Malformative DSD. **(A and B)** Congenital malformations of the cloacal derivatives hampering the identification of the external genitalia: bladder exstrophy in a female newborn (A), and cloacal exstrophy in a male newborn (B). Note the difficulty in distinguishing the derivatives of the labioscrotal swellings (labia majora and scrotum), urethral foldings, and genital tubercle. **(C)** Isolated hypospadias in a boy with normal scrotal configuration. **(D)** In boys with isolated hypospadias, testis hormone secretion (eg, AMH) is normal, as opposed to what is seen in boys with hypospadias because of testicular dysgenesis. AMH levels are represented as standard deviation scores (SDS), where -2 to +2 represent the normal values independent of age. **(A and B:** Reproduced with permission from Ebert AK, Reutter H, Ludwig M, et al. The exstrophy-epispadias complex. *Orphanet J Rare Dis.* 2009 Oct 30;4:23. **C:** Reproduced with permission from Yucel S, Dravis C, Garcia N, et al. Hypospadias and anorectal malformations mediated by Eph/ephrin signaling. *J Pediatr Urol.* 2007 Oct;3(5):354-363. **D:** Reproduced with permission from Grinspon RP, Rey RA. When hormone defects cannot explain it: malformative disorders of sex development. *Birth Defects Res C Embryo Today.* 2014 Dec;102(4):359-373.)

Congenital isolated aphallia has been described in patients with Robinow syndrome (OMIM 268310). **Isolated hypospadias** may be present in the Pallister-Hall syndrome (OMIM 146510), in the X-linked Opitz G/BBB syndrome (OMIM 145410), and in Axenfeld-Rieger syndrome type 1 (OMIM 180500).

Dysgenetic DSD: Abnormal Gonadal Differentiation

Gonadal dysgenesis represents the most conspicuous form of **fetal-onset primary hypogonadism** in which all cell lineages—Leydig/theca, Sertoli/granulosa, and germ cells—are affected. Gonadal dysgenesis may be associated with a diversity of karyotypes: 46,XY, 46,XX, sex chromosome aneuploidies like 45,X and mosaicisms including XY, XX, and/or sex chromosome monosomies or polysomies (see Table 14–2). As suggested by Jost's experiments, defective ovarian development resulting in primary ovarian failure does not affect fetal genital development, so these conditions will not be dealt with in this chapter. Conversely, any disorder affecting the normal process and/or timing of testicular development results in dysgenetic DSD. The severity of the condition may range from complete to very mild, and it may affect the right and left gonads equally, or may be asymmetrical. Finally, a peculiar form of gonadal dysgenesis consists in the development of both ovarian and testicular tissue. The degree of Müllerian duct regression, Wolffian duct development and virilization of the external genitalia depends on the amount of functional testicular tissue secreting androgens (Leydig cells) and AMH (Sertoli cells).

Some sex chromosome aneuploidies, like those observed in Klinefelter syndrome (47,XXY) and Turner syndrome (45,X), will be briefly described by reason of their high prevalence even though they rarely present with malformations of the external or internal genitalia in the newborn.

A. Dysgenetic DSD in 46,XY patients   Gonadal dysgenesis may occur in individuals with a Y chromosome as a consequence of defects in the testis determination pathway explained earlier (Figure 14–8). *SRY* seems to be the only testis-specific gene of the whole pathway, and thus, its mutations result in isolated gonadal dysgenesis, whereas mutations in other genes also result in malformations of non-reproductive organs (see Table 14–2).

Clinical features   Dysgenesis may be **complete**, usually called "pure gonadal dysgenesis" or "agenesis," represented by a streak of fibrous tissue, or **partial**, resulting in a wide spectrum of "testicular dysgenesis." In the latter case, gonads are small, the albuginea is thin, and the seminiferous tubules usually have annular aspect and are separated by wide intertubular spaces. Germ cells are the most affected population, followed by Sertoli cells. These gonads are at a higher risk of development of gonadal tumors, like gonadoblastoma from childhood and seminoma usually after puberty.

In complete or pure gonadal dysgenesis, the fetus is completely feminized, owing to the lack of the two testicular hormones involved in fetal sex differentiation. The diagnosis is usually not suspected until the age of puberty, when these girls present with

46,XY dysgenetic DSD
female/ambiguous genitalia with uterus
low/undetectable T and AMH
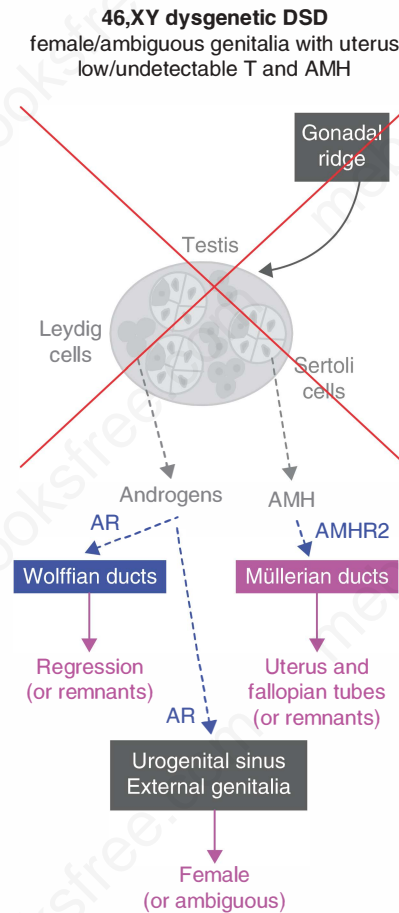
FIGURE 14–8   Pathogenesis of dysgenetic DSD in 46,XY patients AMH, anti-Müllerian hormone; AMHR2, AMH receptor type 2; AR, androgen receptor; DSD, disorder of sex development; T, testosterone.

absence of thelarche and menarche. Sex steroids are extremely low or undetectable, and AMH is undetectable. Gonadotropins are elevated in the newborn and infant, but may decrease to almost normal levels by the age of 6 to 9 years, and increase again to extremely high levels thereafter. In some cases, gonadoblastoma can secrete steroids and result in some breast development.

SRY mutations   (OMIM 480000) Inactivating mutations or deletions of SRY are present in approximately 15% of 46,XY females with complete gonadal dysgenesis (Swyer syndrome). Nonreproductive organs are not affected, so these individuals may seek medical attention only at the age of puberty for the lack of pubertal signs.

SOX9 haploinsufficiency   (OMIM 608160) It, due to heterozygous mutations or deletions, leads to gonadal dysgenesis and campomelic dysplasia (congenital bowing and angulation of long bones), which can also be associated with macrocephaly, micrognathia, low-set ears, flat nasal bridge, congenital dislocation of hips, hypoplastic scapula, small thoracic cage, and cardiac and renal defects. Most patients die during the neonatal period as the result of respiratory distress, although occasionally long-term survival has been reported.

SF1 mutations   Mutations of *NR5A1*, encoding SF1 (OMIM 612965), have been identified in a wide spectrum of disorders of gonadal and adrenal function. However, adrenal function may be normal, and patients present with ambiguous genitalia, regressed Müllerian ducts and biological signs of partial gonadal dysgenesis. Severe cases with complete sex reversal or the opposite, mild forms limited to hypospadias or male factor infertility, have also been described. In mild cases, patients may be transiently fertile and pass the mutation on to the next generation.

WT1 mutations and deletions   (OMIM 607102) In patients with chromosomal aberrations or mutations affecting WT1 expression, gonadal dysgenesis is usually associated with progressive renal disease such as diffuse mesangial sclerosis. Three distinct clinical presentations are recognized. (a) Denys-Drash syndrome is characterized by the clinical triad of gonadal dysgenesis, nephropathy (proteinuria at an early age, evolving into nephrotic syndrome and renal failure owing to focal or diffuse mesangial sclerosis), and Wilms tumor. (b) Frasier syndrome is defined by the coexistence of gonadal dysgenesis, progressive nephropathy and gonadoblastoma, caused by mutations in the donor splice site in intron 9 that results in the +KTS isoform of *WT1*, named after the insertion of the amino acids lysine (K), threonine (T), and serine (S) in the protein. (c) WAGR syndrome is characterized by the association of Wilms tumor, aniridia, genitourinary abnormalities (ambiguous genitalia and/or gonadoblastoma), and mental retardation.

Chromosome 9p deletions   (OMIM 602424) These deletions have been described in patients with gonadal dysgenesis of variable severity in association with mental retardation, microcephaly, facial malformations, short stature, and/or digestive or bronchial malformations. *DMRT1* and *DMRT2* have been identified in the critical 9p24 region.

ATRX mutations or deletions   (OMIM 301040) These mutations or deletions are characterized by the coexistence of α-thalassemia, mental retardation with facial dysmorphism (telecanthus, epicanthus, midface hypoplasia, and small triangular nose with flat nasal bridge and anteverted nares), and ambiguous genitalia.

DHH mutations   These mutations result in 46,XY partial or complete gonadal dysgenesis associated with minifascicular neuropathy (OMIM 605423).

MAP3K1 mutations   (OMIM 600982) These mutations resulting in increased binding to AXIN1 enhance β-catenin expression and reduce SOX9/FGF9 expression. This shifts the balance between the testicular and ovarian pathways, resulting gonadal dysgenesis in 46,XY patients.

DSS duplications   In 46,XY patients with a partial duplication of Xp21.3-p21.2, complete or partial gonadal dysgenesis results in female or ambiguous genitalia, respectively. The locus was named DSS, for dosage-sensitive sex-reversal (OMIM 300018), because a double active dose (ie, escaping X inactivation) is likely the underlying cause of gonadal dysgenesis. This is not observed in patients with a complete X polysomy (eg, Klinefelter syndrome) because one of the X chromosomes undergoes inactivation. The smallest duplication includes *NR0B1*, which encodes **DAX1**, suspected to be the underlying cause; however, other contiguous genes also exist in the duplicated region and their implication cannot be excluded.

Environmental causes may underlie the increase in the incidence of hypospadias, cryptorchidism, low sperm count, and testicular cancer. It has been proposed that these signs are all part of the same underlying entity, named testicular dysgenesis syndrome. Cryptorchidism and other genital malformations have been observed with increased incidence in areas of intensive agriculture and in children of workers exposed to pesticides with estrogen-like action.

B. Dysgenetic DSD in patients with mosaicisms carrying a Y chromosome, including ovotesticular DSD and asymmetric gonadal differentiation   In patients with mosaicisms including a Y chromosome, testicular tissue usually develops in a dysgenetic context. When an XX cell lineage is also present, ovarian tissue may develop as well (see Table 14–2).

Ovotesticular DSD is reserved for the existence of both testicular tissue, including seminiferous tubules, and ovarian tissue, where the presence of follicular structures containing oocytes is mandatory; the sole existence of an "ovarian-like" stroma is insufficient. Ovotesticular DSD may present with bilateral ovotestes—gonads carrying both testicular and ovarian tissue—one testis and one ovary or one ovotestis and one testis or ovary. The testicular tissue is usually dysgenetic, whereas the ovarian tissue is most frequently of normal appearance. The occurrence of a 46,XX/46,XY karyotype or mosaicisms carrying both an XX and an XY lineage explains the existence of ovarian and testicular tissues. Ovotesticular DSD can also present in 46,XX individuals (see later) but is extremely rare in 46,XY patients.

The overall incidence of ovotesticular DSD is reported at 3% to 10% of cases of DSD. Ovotesticular disorder shows the highest prevalence among South African blacks with the number of published cases being 17 per 100 million people, followed by Europe at 15.3, and Asia at 1.2 cases per 100 million. The age and clinical features at presentation vary according to the amount of functional testicular and ovarian tissues. Patients may be almost completely virilized, with mild signs of androgen insufficiency like hypospadias and undescended testes. Serum androgen and AMH levels are usually within the male range, and gonadotropins are not elevated. At the age of puberty, exaggerated follicular development may provoke abdominal pain; estradiol may increase and result in breast development, and in extreme cases cyclic urethral bleeding may indicate the existence of Müllerian remnants. When the ovarian tissue predominates, and there is scarce testicular tissue, the newborn is feminized, with clitoromegaly and labial fusion. Androgens and AMH are elevated for a girl but low for a boy, and gonadotropins are usually within the normal female range.

Müllerian derivatives are usually present in less virilized patients; however, ultrasonographic assessment may be deceiving as a diagnostic tool in these newborns or infants due to difficulties in imaging resolution. The risk of gonadal malignancy is low, roughly between 2% and 5%, although it may vary according to the presence or absence of a Y chromosome.

Asymmetric gonadal differentiation, also known as **mixed gonadal dysgenesis**, is characterized by the presence of a more or less dysgenetic testis on one side and a streak on the other. The most frequent karyotype is 45,X/46,XY, where the predominance of the 45,X on one side explains the development of the streak and the predominance of the 46,XY lineage underlies the development of the testis on the other. However, a 45,X/46,XY karyotype should not be equated to asymmetric gonadal differentiation (see Table 14–2). Indeed, this karyotype can be an incidental finding in a normally virilized boy, and also asymmetric gonadal differentiation may occur in a 46,XY patient. The diagnosis is histological.

Clinically, the patient usually presents with asymmetric development of the scrotum, with the testis usually being in scrotal position. The internal ducts, as expected, are typically male on the side of the testis and female (hemi-uterus and Fallopian tube) on the side of the streak. Some patients may also have features typical of Turner Syndrome. Serum levels of androgens and AMH correlate with the amount of functional testicular tissue. The risk of gonadal tumor development is high.

C. Dysgenetic DSD in 46,XX patients, including ovotesticular DSD and testicular DSD

Testicular tissue can differentiate in the gonadal primordium of 46,XX fetuses, either as the only gonadal tissue in a condition called **46,XX Testicular DSD**, or together with ovarian tissue in **46,XX Ovotesticular DSD**. Both DSD forms can be further categorized as SRY-positive and SRY-negative groups (see Table 14–2). The existence of the SRY sequence explains the development of testicular tissue. In most instances, the *SRY* gene has been translocated to an X chromosome, due to recombination between the pseudoautosomal regions of the X and Y chromosomes. However, translocation to an autosome can occur. One example was the incidental finding of small testes, azoospermia, and translocation of *SRY* onto the terminal end of chromosome 16q in a 61-year-old 46,XX man. Indeed, *SRY*-positive 46,XX individuals most frequently have testes with only mild dysgenetic features (eg, scarce germ cells) and present as 46,XX males. Conversely, *SRY*-negative patients most frequently have dysgenetic testes or ovotestes and are diagnosed at birth because of ambiguous genitalia.

Molecular etiologies have been identified for some *SRY*-negative 46,XX individuals with testicular or ovotesticular DSD. They are associated with overexpression of "pro-testicular" genes, like *SOX9*, *SOX10*, and *SOX3*, or defects of "pro-ovarian" genes, like *RSPO1* and *WNT4* (see Figure 14–3).

SOX9, SOX10, and SOX3
Overexpression due to duplications of the *SOX9* gene or of non-coding regulatory elements upstream of *SOX9* have been reported in SRY-negative sex-reversed 46,XX individuals. These regulatory elements appear to include 40.7-41.9 kb sequences located approximately 600 kb upstream of *SOX9*.

Another association reported in 46,XX *SRY*-negative male is overexpression of the *SOX10* gene, which is located at 22q13. Gain of function *SOX3* mutations, located at Xq27, has also been associated with XX male sex reversal.

RSPO1
Inactivating mutations in *RSPO1* gene, located at chromosome 1p34, are associated with 46,XX sex reversal. In addition, affected individuals have palmoplantar hyperkeratosis with squamous cell carcinoma of skin and lack Müllerian structures.

WNT4
In a consanguineous family, fetuses and infants homozygous for a loss-of-function mutation in *WNT4*, located at chromosome 1p36.2, showed sex reversal and aberrant development of kidneys, adrenal and lungs. This compilation of features has been labeled as SERKAL syndrome. Heterozygous mutations in *WNT4* have been detected in 46,XX women with primary amenorrhea associated with Müllerian duct abnormalities and androgen excess but no existence of testicular tissue.

Clinical features
Patients with testicular DSD that are completely virilized correspond to the original description of "XX male," as defined by de la Chapelle. That definition was an individual of *"a male phenotype, male psychosexual identification, testes without macroscopic or microscopic evidence of ovarian tissue, and absence of female genital organs."* Approximately 90% of these XX males are *SRY*-positive. In these typical cases, diagnosis is usually delayed until adult life when the patient seeks medical care for infertility. Less frequently, boys may be studied for short stature or small testes at the age of puberty. In fact, XX boys are shorter than age-matched XY pairs, in spite of similar circulating levels of hormone of the growth axis, including IGF1 and IGFBP3. The existence of Y-specific growth genes may be the explanation. Testicular volume does not differ from that of control boys during childhood. The underlying reason is that testicular volume depends mainly on Sertoli cells before puberty. Sertoli and Leydig cell function is normal in childhood and in the early stages of puberty, as revealed by normal AMH, inhibin B, testosterone and gonadotropin levels. Like in Klinefelter syndrome, testicular volume increases progressively to 4 mL, and even to 6 to 8 mL mainly due to Sertoli cell proliferation, but then remains smaller due to germ cell depletion. The existence of two X chromosome and the absence of a Y chromosome results in meiotic failure, azoospermia and low semen volume. Seminiferous tubule dysfunction in the advanced stages of puberty and adulthood is reflected in low inhibin B and increasing FSH, whereas a mild Leydig cell dysfunction may exist in some cases leading to low normal testosterone and increased LH. A typical feature is the elevation of estrogen levels responsible for gynecomastia.

Testicular DSD with more severe dysgenetic features, usually associated with SRY-negative cases, present with a similar clinical spectrum to that described earlier for dysgenetic DSD in 46,XY patients. Clinical features of patients with 46,XX ovotesticular

DSD is similar to that described in the previous section; indeed, the presentation does not depend on the karyotype.

D. Gonadal dysgenesis in sex chromosome aneuploidies

Atypical sex chromosome constitution includes polysomies of the X chromosome in males (Klinefelter syndrome) and females (Triple X syndrome) and monosomy of the sex chromosome pair (45,X), with their mosaic variants. Usually, these disorders do not represent a differential diagnosis in patients with ambiguous genitalia. In the case of X chromosome polysomies, the gonadal dysgenesis is mild and of late onset, thus explaining why fetal development is not affected. However, they will be briefly described here because they conform to the classification of Sex-Chromosome DSD.

Klinefelter syndrome is the most common cause of male hypogonadism, with a prevalence of approximately 1.5 in 1000 males. Its first description in 1942 was clinical, and included eunuchoidism, gynecomastia, small testes, azoospermia, and elevated FSH. In 1959, the condition was associated to a 47,XXY karyotype. Over the past decades, the clinical spectrum of the syndrome has expanded to include patients with pure cell lines or mosaicisms with two or more X chromosomes and a Y chromosome. However, some authors warn that patients with 48,XXYY, 48,XXXY, and 49,XXXXY polysomies present more complex physical, medical, and psychological phenotypes and should not be considered as Klinefelter syndrome even though they share features of primary hypogonadism.

Gonadal dysgenesis is extremely mild during fetal life, only characterized by a decrease in germ cell numbers, but with no impairment of Leydig and Sertoli cell function until midpuberty. Therefore, internal and external genitalia are normally virilized in Klinefelter patients. Generally, at birth, no particular stigmata occur in Klinefelter syndrome. During childhood, no specific signs or symptoms are noted, except for an increased risk of cryptorchidism, learning disabilities, or behavior disorders. Gonadotropins, testosterone, AMH and inhibin B are within the normal range. Klinefelter boys are long-legged, most probably as the result of an overdose of the X-chromosome *SHOX* gene.

Pubertal development is usually spontaneous, but characterized by the lack of progression of testicular size beyond 6 to 8 mL and the development of gynecomastia. Due to meiotic failure associated with X chromosome polysomy, the seminiferous tubules do not enlarge, but rather undergo fibrosis and hyalinization, which results in small firm testes and azoospermia. Nonetheless, microscopic areas of the testes may be normal, and spermatozoa can be recovered by testicular puncture. The marked reduction of tubular volume accounts for the apparent hyperplasia of the interstitial tissue. Leydig cell function becomes progressively deficient, resulting in a reduced testosterone production rate with compensatory LH hypersecretion. Because of the high LH levels, estradiol secretion is stimulated, with an increased estradiol/testosterone ratio, which probably accounts for the development of gynecomastia.

47,XYY males and 47,XXX females are almost as frequent as Klinefelter syndrome, with a rough prevalence of 1 in 1000 live births. However, they are rarely diagnosed because they are usually oligo- or asymptomatic. Gonadal development and function does not show major impairment, except for premature ovarian failure described in XXX women. Like Klinefelter patients, these children are long-legged and might show behavior disturbances.

Turner syndrome was also initially a clinical description of a distinct entity associating sexual infantilism, webbed neck, and cubitus valgus in females. Definitive evidence for a 45,X karyotype was provided several decades later. Turner syndrome occurs in approximately 1 in 2500 live female births, but the number of unborn affected fetuses is much larger, since spontaneous abortion occurs in approximately 98% of pregnancies.

The syndrome spans a wide spectrum of clinical presentations. The main features include gonadal dysgenesis; short stature; a short, thick, and webbed neck; hypertelorism; epicanthus and ptosis of the eyelids; low posterior hairline; congenital lymphedema; broad child-like chest; cubitus valgus and short fourth metacarpals; multiple pigmented nevi; intestinal telangiectasia; recurrent otitis media; and a tendency to keloid formation. There is an increased risk of celiac disease, autoimmune thyroiditis, diabetes mellitus, rheumatoid arthritis, inflammatory bowel disease, cardiovascular defects—especially coarctation of the aorta and nonstenotic bicuspid aorta—and kidney anomalies. Bone age is delayed, and bone mineral density may be affected. Newborns are smaller than average both in length and in weight. Usually, postnatal growth lies in the normal range for the first 2 to 3 years, but then decreases, and height deviates progressively from the norm.

Ovarian follicles undergo premature apoptosis resulting in fibrous streaks as typical histologic findings. Gonadotropin levels are high during infancy, but may decrease to normal levels between 6 and 9 years of age, whereas undetectable AMH throughout childhood is indicative of the absence of ovarian tissue and predicts the likely absence of spontaneous pubertal development.

Turner syndrome variants with partial gonadal dysgenesis may occur in patients with mosaicisms (eg, 45,X/46,XX or other variants) or in cases of partial X chromosome losses (eg, isochromosomes of the long or the short arms, ring chromosomes). The normal X chromosome comes from the mother in most patients with a 45,X karyotype, but only in about half of those with isochromosomes or ring chromosomes. Turner syndrome variants with partial gonadal dysgenesis may present with less anatomic stigmata and spontaneous pubertal onset; however, primary or secondary amenorrhea, due to premature ovarian failure is the rule. Detectable, or even normal serum AMH, indicating the existence of ovarian follicles, may be predictive of spontaneous puberty. In these patients, gonadotropin levels may be only moderately elevated. Spontaneous pregnancy has been described but is an exceptional event.

Non-dysgenetic DSD with Testicular Differentiation

Even if testicular differentiation occurs following the expected early pathway in the 46,XY fetus, insufficient virilization or feminization of the genitalia may result when one of the two hormone-driven processes of genital development—the androgen-dependent or the AMH-dependent—are defective (see Table 14–2). As can be
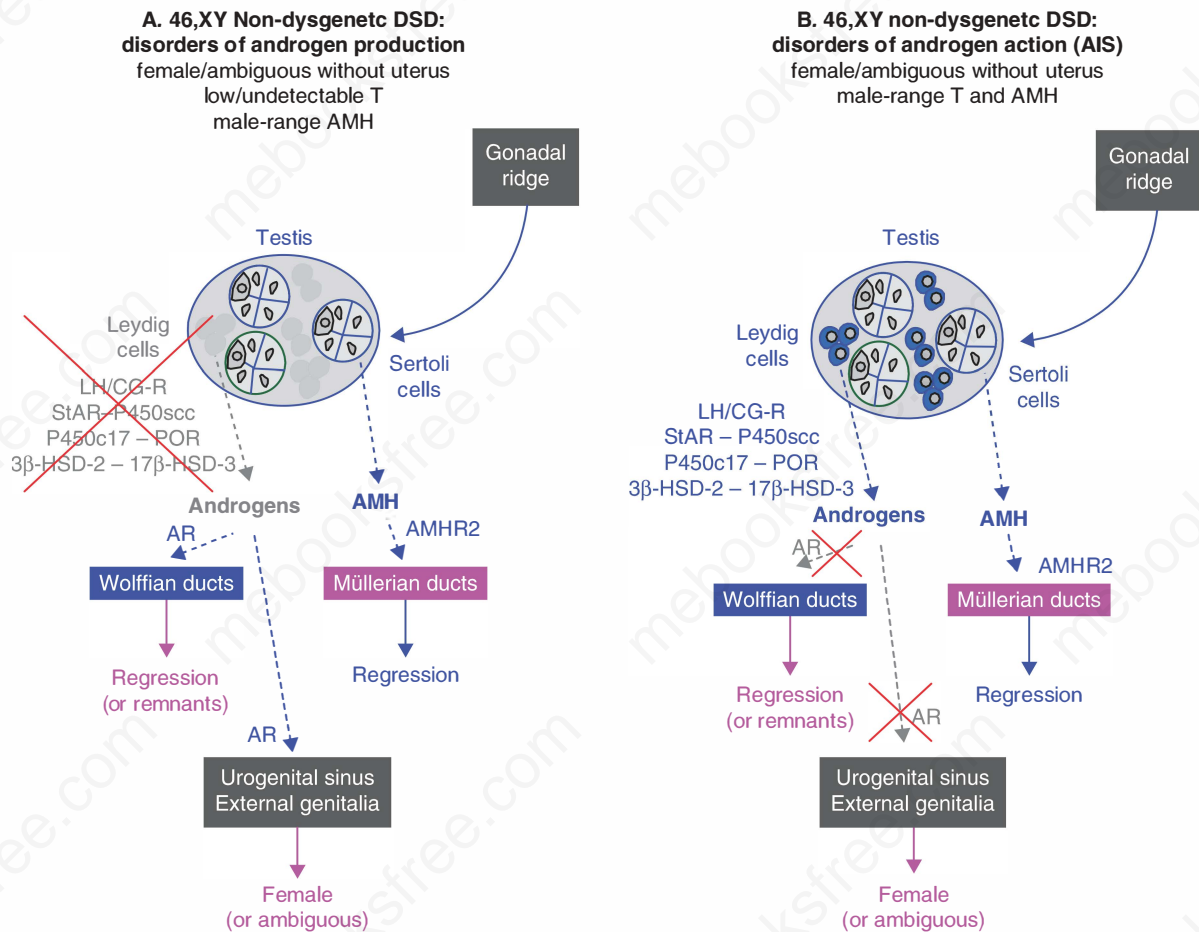
FIGURE 14–9 Pathogenesis of non-dysgenetic DSD in 46,XY patients. **(A)** Defects in the production of androgens. **(B)** Androgen insensitivity syndrome (AIS). 3β-HSD-2, 3β-hydroxysteroid dehydrogenase type 2; 17β-HSD-3, 17β-hydroxysteroid dehydrogenase type 3; AMH, Anti-Müllerian hormone; AMHR2, AMH receptor type 2; AR, Androgen receptor; DSD, Disorder of sex development; LH/CG-R, Luteinizing hormone/Chorionic gonadotropin receptor; P450c17, Cytochrome P450 17-hydroxylase/17,20-lyase; P450scc, Cytochrome P450 side-chain cleavage; POR, Cytochrome P450 oxidoreductase; StAR, Steroidogenic Acute Regulatory protein; T, Testosterone.

expected, the phenotypes are entirely opposed according to which is the deficient pathway: whereas defects in androgen synthesis or action drive to a female without a uterus (Figure 14–9), defects in the AMH pathway lead to a male with a uterus.

A. Disorders of androgen synthesis In contrast to testicular dysgenesis, characterized by an early onset fetal hypogonadism, with whole testicular dysfunction, defects of androgen synthesis represent one form of "dissociated" or "cell-specific" fetal onset primary hypogonadism, in which Leydig cells, but not Sertoli cells, are primarily affected.

Defects in LHCG Receptor **Leydig Cell Aplasia or Hypoplasia** (OMIM 238320) is a rare autosomal recessive form of DSD due to inactivating mutations of the LH/chorionic gonadotropin (CG) receptor (LH/CGR). Leydig cell differentiation and function are impaired or completely abolished, thus resulting in a partial or total deficiency of androgen production by the testes. The phenotypic spectrum ranges from female external

genitalia to mild undervirilization. The karyotype is 46,XY, serum AMH is within the normal male range, and Müllerian derivatives are absent. Basal and human (h) CG-stimulated testosterone is low or undetectable, plasma LH is elevated, and FSH is normal. Testes are inguinal or intraabdominal and contain seminiferous tubules with Sertoli and germ cells but no Leydig cells. Most of the mutations found in the *LH/CGR* gene, located at chromosome 2p16.3, are in exon 11, encoding the seven transmembrane domains.

Defect of Cholesterol Biosynthesis **Smith-Lemli-Opitz Syndrome** (SLOS, OMIM 270400) is due to a dehydrocholesterol reductase (DHCR) defect that blocks the last step of cholesterol synthesis. Since cholesterol is also required for signal transduction of SHH in the development of the nervous system, face and limbs, undervirilization in 46,XY patients with SLOS is associated with microcephaly, low set ears, short and broad nose, cleft palate, broad gingival ridges, growth retardation, toe syndactyly, pyloric stenosis and congenital heart defects.

communities to change their social sex from female to male. At variance with the other androgen synthesis defects, sperm production is feasible when testes are brought to scrotal position. Serum hormone levels are characterized by a high T/DHT ratio (usually above 18-20) after hCG stimulation. However, the diagnosis is more accurate when an elevated $5\beta/5\alpha$ urinary metabolite ratio is found. Over 50 different mutations have been recorded in *SRD5A2*, which maps to chromosome 2p23.1. The genotype/phenotype correlation is poor.

Recently defects in the **"backdoor" pathway** have been identified in patients with deficient DHT production. Inheritance is also autosomal recessive, due to mutations in *AKR1C2* (OMIM 600450), encoding 3α-HSD type 3, and in *AKR1C4* (OMIM 600451), encoding 3α-HSD type 1. Both genes map to mapping to 10p15.1 (see Figure 14–5).

Other Defects of Androgen Biosynthesis   Mutations in *MAMLD1* (OMIM 300120) are associated with hypospadias and other defects of genital virilization, probably due to defects in testosterone production. The precise mechanisms by which MAMLD1 regulates steroid synthesis needs to be fully clarified.

B. Disorders of androgen action: The Androgen Insensitivity Syndrome (AIS)
More than 800 mutations have been recorded in the *AR* gene, mapping to Xq12, in patients with AIS (http://androgendb.mcgill.ca).

Complete Androgen Insensitivity Syndrome (CAIS)   Mutations that completely abolish the AR activity by affecting the formation of the steroid receptor-coregulator complex, its nuclear transport, its binding to DNA, or its transcriptional activation ability, underlie CAIS. It is the most frequent cause of DSD in completely feminized 46,XY subjects. Patients appear as normal females at birth and come to medical attention at puberty because of primary amenorrhea, after normal breast development due to estrogen synthesis from testicular testosterone production. Alternatively, the diagnosis may be made in girls because of an inguinal hernia containing a palpable gonad that turns out to be a testis, or prenatally because of discordance between genotype and phenotype. Pubic and axillary hair is absent, the vagina is short with a blind ending, and the uterus is absent, as a consequence of normal AMH action in fetal life (see Figure 14–9). The testes appear quite normal at birth, but germ cell number rapidly declines; due to androgen insensitivity, meiosis cannot proceed and spermatozoa are not produced. The risk of germ cell malignant degeneration is low, but benign tumors such as Sertoli cell adenomas or hamartomas are more frequent. Laboratory findings are characterized in non-gonadectomized patients by serum testosterone in the male range with elevated gonadotropins at pubertal age, but they can be low in the neonatal period. AMH levels are in the male range.

Partial Androgen Insensitivity Syndrome (PAIS)   is defined by some degree of virilization, which may range from predominantly female genitalia to phenotypic males with hypospadias and micropenis. The diagnosis is difficult in the absence of family history and cannot easily be distinguished from that of DHT synthesis defects. Hormone levels are usually in the male range, including testosterone, gonadotropins, and AMH, which distinguishes PAIS from gonadal dysgenesis in 46,XY DSD. Response to hCG stimulation is good in terms of testosterone production; this feature helps eliminate androgen synthetic defects from the differential diagnosis. At puberty, scanty pubic and axillary hair develop; gynecomastia is frequent; testosterone increases; and AMH may decrease in correlation with the degree of retained androgen action. There is poor genotype/phenotype correlation, and the same mutation can result in widely different phenotypes, even within the same family.

C. Disorders of AMH synthesis or action
The **persistent Müllerian duct syndrome (PMDS)** (OMIM 261550) is a rare, autosomal recessive form of DSD characterized by persistence of the uterus and Fallopian tubes in a boy otherwise normally virilized (Figure 14–10). The existence of other signs of undervirilization, like hypospadias, should be ruled out in order to distinguish this condition from gonadal dysgenesis. The presence of Müllerian derivatives is usually discovered at surgery for cryptorchidism or inguinal hernia. Two anatomical forms of PMDS have been described: in 1/5 of cases, the patient presents with bilateral cryptorchidism, and both testes are embedded in the broad ligament in an "ovarian" position, whereas in the more frequent presentation one testis has descended into the scrotum, dragging the Fallopian tube and the contralateral testis. Mutations in the AMH or the AMH receptor type 2 (AMHR2) receptor genes are found in approximately 85% of PMDS patients. Transmission is autosomal recessive. Idiopathic PMDS, not due to gene mutations, is frequently associated with associated abnormalities. Serum testosterone and gonadotropins are in the male range. AMH is undetectable in patients with *AMH* mutations, but within the male range in patients with *AMHR2* mutations. The risk of testicular tumors is not higher than in other patients with longstanding cryptorchidism. The main goal of treatment is to place the testes in the scrotum when necessary. If the uterus and Fallopian tubes limit mobility, proximal salpingectomy and excision of the uterine fundus may be performed after meticulous dissection of the vasa deferentia, leaving the cervix and pedicles of the myometrium in place to preserve vascularization.

Non-dysgenetic DSD with Ovarian Differentiation

Exposure to elevated androgen levels during intrauterine life underlies virilization of the 46,XX fetuses in whom ovarian differentiation occurred (see Table 14–1). Individuals with 46,XX DSD and bilateral ovaries come to medical attention because of genital ambiguity at birth or postnatal virilization during childhood or at the time of expected puberty. As noted previously, the external genital development of 46,XY and 46,XX fetuses is similar until 6 to 7 weeks of gestation. The typical female fetus does not secrete androgens during gestation. Exposure of female fetuses to androgens during gestation alters the developmental
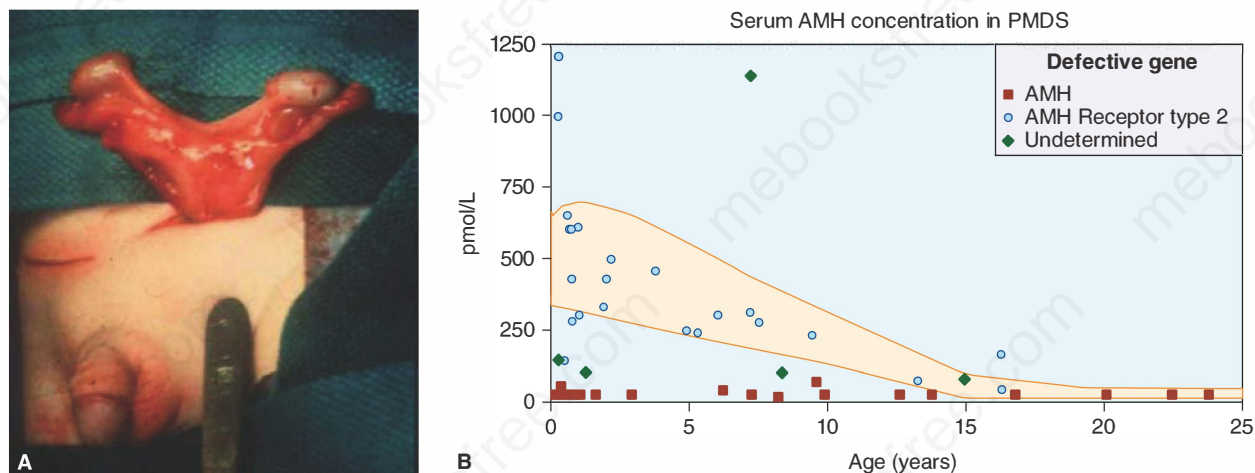
FIGURE 14–10 **(A)** Persistent Müllerian duct syndrome (PMDS) characterized by the persistence of uterus and Fallopian tubes in an otherwise normally virilized boy. **(B)** Patients with undetectable AMH have AMH gene mutations (red); those with normal or high AMH usually have AMH receptor type 2 mutations (blue), but in a few cases no mutation has been detected (undetermined, green). (Reproduced with permission from Jameson JL, De Groot LJ, de Kretser D, et al. *Endocrinology: Adult and Pediatric*, 7th ed. Philadelphia, PA: Elsevier Saunders; 2015.)

programming and can lead to masculinization of the external genitalia. However, since AMH is not produced during the period of Müllerian duct sensitivity, the uterus and Fallopian tubes develop. Excessive androgen sources can be fetal (eg, the virilizing form of CAH, which is the most common cause of genital ambiguity in a 46,XX infant), but also maternal (eg, steroid secreting tumors) or placental (eg, aromatase deficiency).

A. Disorders of adrenal steroidogenesis   Inheritance of these disorders of steroidogenesis is autosomal recessive. In **CAH**, cortisol biosynthesis is impaired due to mutations in genes encoding specific steroidogenic enzymes. Insufficient cortisol production leads to loss of negative feedback inhibition at the pituitary, increased ACTH secretion, and buildup of adrenal steroid hormones proximal to the deficient enzyme. Clinical features and laboratory data vary depending on which enzyme gene is involved. The most common type of CAH is 21-hydroxylase deficiency (21-OHD, OMIM 613815). The specific mutations and residual enzyme activity influence the extent of glucocorticoid and mineralocorticoid hormone deficiencies.

Accumulation of steroid hormone intermediates such as 17-hydroxyprogesterone (17-OHP) can drive adrenal steroid hormone metabolism toward the so-called alternate or "backdoor" steroid pathway (see Figure 14–5). In this pathway, excessive 17-OHP undergoes 5α and 3α-reduction to 5α-pregnane-3α,17α-diol-20-one (pdiol), ultimately generating androstanediol which is the substrate for 3α-oxidation and subsequent conversion to DHT bypassing testosterone synthesis. Mutations in 21-hydroxylase (*CYP21A2*), 11β-hydroxylase (*CYP11B1*), or P450-oxidoreductase (*POR*) increase flux through this pathway leading to elevated DHT concentrations and prenatal virilization of 46,XX fetuses. The relevance of other intra-adrenal C19 steroidogenic pathways is becoming apparent. For example, androstenedione

can be converted to 11β-hydroxy-androstenedione, which can be converted to more potent C19 androgens.

Determination of hormone concentrations is essential for the diagnosis of disorders affecting steroidogenesis. Traditionally, hormone concentrations have been measured using radioimmunoassays. However, radioimmunoassays can be problematic due to cross-reactivity with similar steroid compounds and limited sensitivity. The newer commercial direct platform immunoassays have introduced additional disadvantages including greater interference with steroid binding proteins and increased non-specific competition with the antibodies used in the assay. The use of liquid chromatography-tandem mass spectrometry (LC-MS/MS) enables simultaneous measurement of multiple steroids and can improve diagnostic testing for disorders of steroidogenesis. Urinary GC-MS avoids the trauma of collecting blood. Another advantage offered by urinary GC-MS is urinary steroid profiling which can lead to better characterization of disorders of steroidogenesis.

21-Hydroxylase Deficiency (21-OHD)   The most common type of CAH (accounting for 90%-95% of cases) is 21-OHD, due to mutations in the 21-hydroxylase (*CYP21A2*) gene located at chromosome 6p21.3 in the HLA class III region. This enzyme converts 17-OHP to 11-deoxycortisol (S) (see Figure 14–5). The reported incidence of 21-OHD ranges from 1 in 9000 to 1 in 15,000 for the **classic form** and 1 in 1000 for the **nonclassic form (NCAH)**. The classic form is categorized into **salt-losing** and **non-salting losing** forms. Estimates of *in vitro* 21-hydroxylase activity are less than 1% for mutations associated with salt-losing CAH, 2% to 10% for simple virilizing CAH, and 30% to 50% for NCAH. This classification system is somewhat contrived because disease severity is better depicted as a continuum based on residual enzyme activity. Most affected individuals are compound heterozygotes with different mutations on each allele. An affected

individual's phenotype is usually defined by their milder mutation. The frequency of CAH varies among ethnic/racial backgrounds.

The spectrum of impaired 21-hydroxylase activity ranges from complete glucocorticoid and mineralocorticoid deficiencies to mild deficiencies manifested principally by compensatory excessive adrenal androgen secretion. Ages at presentation range from infancy to adulthood depending on the clinical manifestations and severity of the mutations. Affected 46,XX infants with severe complete loss of function mutations typically present in the neonatal period with genital ambiguity. They can develop dehydration, hypoglycemia, hyponatremia, and hyperkalemia when the diagnosis is delayed or missed. If not identified by a newborn screening program, infant boys typically present at 2 to 3 weeks of age with failure to thrive, poor feeding, lethargy, dehydration, hypotension, hyponatremia, hyperkalemia, and hyperpigmentation of the scrotum. Missed or delayed diagnosis can lead to a fatal outcome. In addition to being life threatening, salt-losing crises have been associated with reduced cognitive function. In the absence of prenatal cortisol exposure, adrenomedullary development is aberrant in classic CAH and is often associated with epinephrine deficiency and hypoglycemia.

For affected 46,XX fetuses, the spectrum of prenatal external genital virilization ranges from clitoromegaly to perineal hypospadias to complete fusion of urethral and labioscrotal folds giving rise to a phallus with chordee and urethral meatus at the tip of phallus and a fused empty scrotum. External genital virilization may be so extensive that affected female infants appear to be males with bilateral undescended testes (Figures 14–11 and 14–12). Even in the absence of overt salt loss due to mineralocorticoid deficiency, there may be sufficient prenatal androgen exposure to virilize the external genitalia. Typically, the urethra and vagina open into the urogenital sinus with a single perineal orifice. For some girls, the urogenital confluence is high in the pelvis with the vagina opening into the urethra just below the bladder. Despite the excessive prenatal androgen exposure, ovarian position and uterine development are normal in affected females. Thus, physical examination showing bilateral undescended testes with a phallus and perineoscrotal hypospadias should prompt an evaluation for a virilizing CAH.

CAH due to 21-OHD is an autosomal recessive disorder characterized by *CYP21A2* mutations on both the maternal and paternal alleles. The gene is located within a complex genomic region at chromosome 6p21.3 and is approximately 30 kb from a highly homologous pseudogene, *CYP21A1P*. These genes are arranged in tandem repeats with the C4A and C4B genes and lie in close proximity to the tenascin (TNX) and serine threonine nuclear protein kinase (RP) genes. The tenascin-XB (*TNXB*) gene codes for an extracellular matrix protein. These four genes, *RP*, *C4*, *CYP21*, and *TNX*, form a unit known as RCCX. Most alleles carry two RCCX units in which one has *CYP21A2* and the other has *CYP21A1P*.
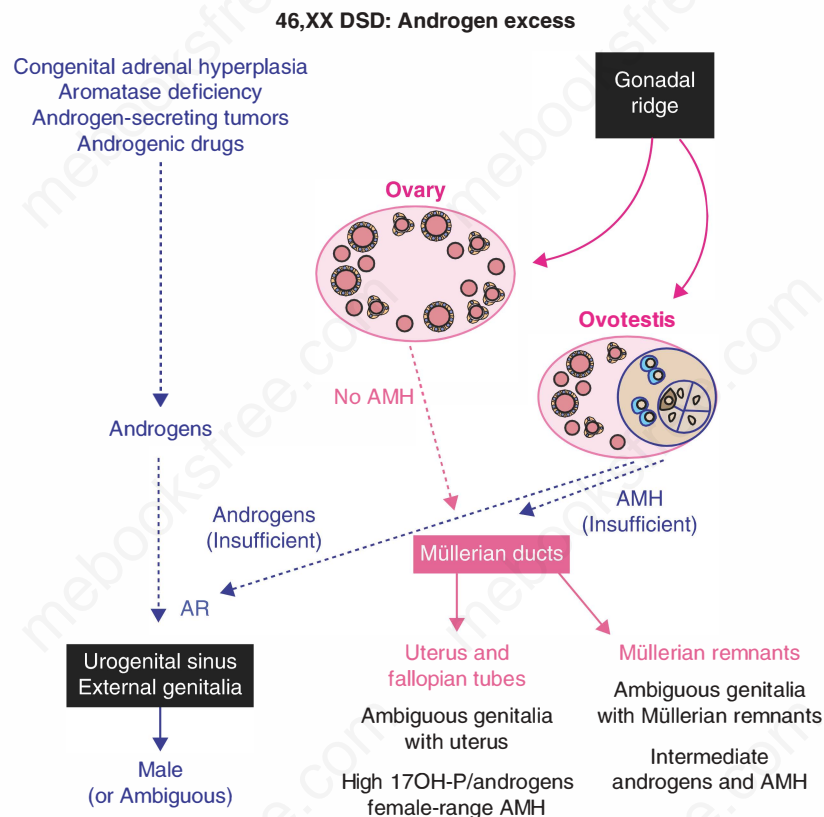
FIGURE 14–11 Pathogenesis of DSD in 46,XX patients. (17OH-P, 17 hydroxyprogesterone; AMH, anti-Müllerian hormone; AR, androgen receptor; DSD, disorder of sex development.)
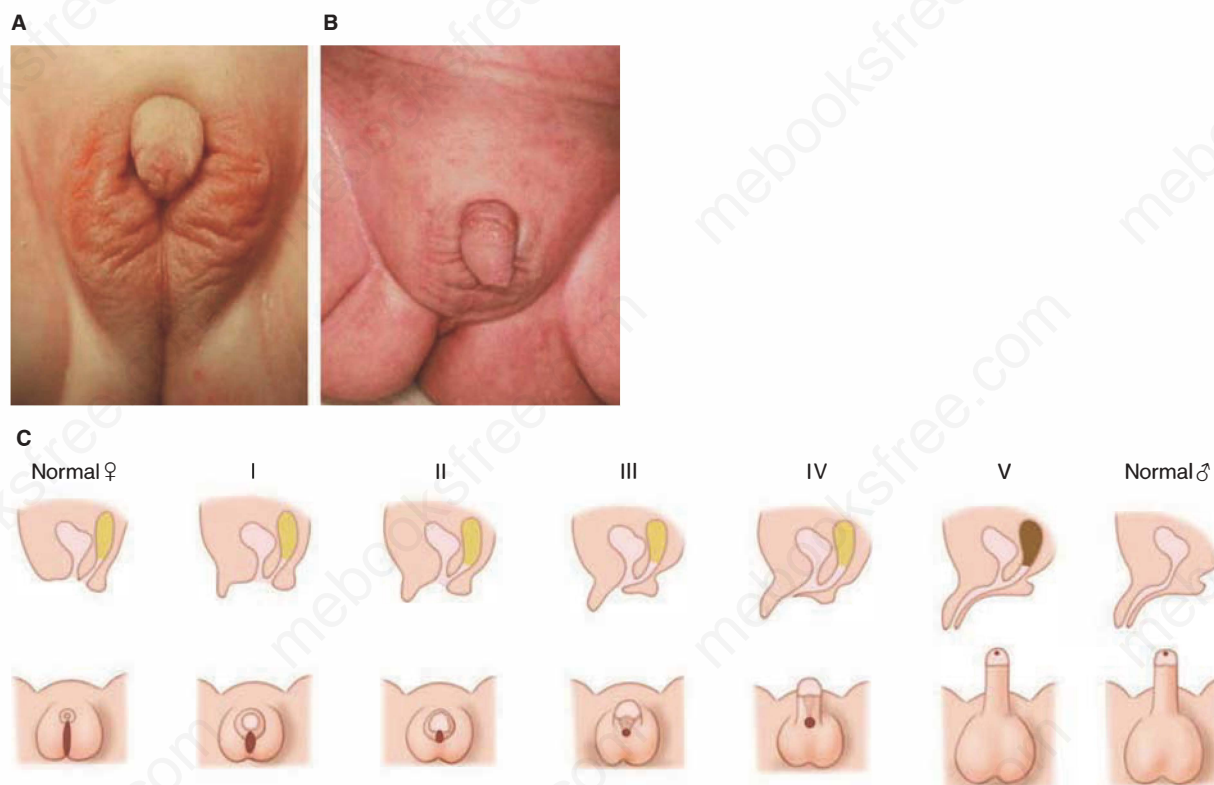
FIGURE 14–12   Genitalia of 46,XX patients with congenital adrenal hyperplasia (CAH) resulting from 21-hydroxylase deficiency. **(A)** Prader stage IV. **(B)** Prader stage V. **(C)** Schematic of Prader staging for patients with CAH. **(A and B:** Reproduced with permission from Jameson JL, De Groot LJ, de Kretser D, et al. *Endocrinology: Adult and Pediatric*, 7th ed. Philadelphia, PA: Elsevier Saunders; 2015. **C:** Adapted with permission from Ogilvy-Stuart AL, Brain CE. Early assessment of ambiguous genitalia. *Arch Dis Child*. 2004 May;89(5):401-407.)

To date, over 150 *CYP21A2* mutations have been reported (www.cypalleles.ki.se and www.hgmd.cf.ac.uk). Most affected individuals are compound heterozygotes. Some alleles carry multiple mutations. In general, phenotype correlates with genotype such that the phenotype reflects the less severe mutation and residual 21-hydroxylase enzyme activity. Nevertheless, a small number of mutations account for the majority of affected alleles. The *de novo* mutation rate is reported to be less than 1%. The most frequently identified mutations represent gene conversion events in which *CYP21A2* has acquired deleterious *CYP21A1P* sequences. Misalignment during meiosis followed by genetic recombination can cause deletions and duplications giving rise to duplications or deletions of the RCCX unit. Haplotypes with 3 or 4 RCCX units have been described. Another example of misalignment is the *CYP21A1P/CYP21A2* chimera. Rarely, uniparental disomy can lead to CAH.

Following a pilot newborn screening program in 1977, most countries have implemented newborn screening programs using filter-paper whole blood 17-OHP levels. The blood sample should be collected after 48 hours of life. Automated time-resolved-dissociation-enhanced lanthanide fluoroimmunoassay (DELFIA) is often used for 17-OHP determinations. The goal of newborn screening has been to identify infants with salt-losing CAH and prevent mis-identification of affected females. Newborn screening

programs have decreased the morbidity and mortality associated with acute adrenal insufficiency. However, preterm and stressed infants can have elevated screening values leading to false positive results. In addition, cross-reactivity with sulfated steroids and 16α-hydroxyprogesterone contributes to false positive results. Decreased 11β-hydroxylase activity in the neonate may be another factor contributing to false positive testing. To avoid excessive recalls for false positive tests, birthweight and gestational age cut-points have been developed. False negative 17-OHP results leading to delayed diagnoses have been reported for both newborn girls and boys.

Discriminating individuals with NCAH from carriers may be challenging. The complexity of the *CYP21A2* locus complicates molecular diagnosis and precludes its use as the first line diagnostic test. At the present time, functional testing such as hormone determinations and/or ACTH stimulation tests may be necessary to confirm decreased 21-hydroxylase activity. Multiple genetic testing strategies such as PCR-based mutation detection methods, sequencing, and multiplex ligation-dependent probe amplification may be needed to accurately ascertain the mutations in an affected individual. Genotyping of parents to assign the affected maternal and paternal mutations and confirm that mutations are on opposite alleles may be essential. Genetic analysis can be a useful adjunct to newborn screening. It is important to acknowledge that this is a complex genetic locus, multiple mutations can occur

on a single allele and that different *CYP21A2* mutations can occur in one family.

Postnatal virilization continues in individuals with simple virilizing CAH who were not identified by newborn screening. Affected children present with premature development of pubic hair, tall stature, accelerated linear growth velocity, acne, and advanced skeletal maturation. Affected prepubertal boys may present with premature pubic hair, phallic enlargement, and bilateral small testes. The missense mutation, I172N, is associated with simple virilizing CAH.

NCAH is the most common form with increased prevalence in Ashkenazi Jews and Hispanics. Individuals with NCAH secrete sufficient cortisol and aldosterone to prevent symptoms of glucocorticoid and mineralocorticoid deficiency in association with excessive C19 adrenal androgen secretion. Clinical features range from premature pubarche to an asymptomatic state. The hyperandrogenic symptoms lead to an ascertainment bias such that more affected women are identified. Premature pubarche is the most common presenting feature in prepubertal children. Hirsutism is the most common presenting feature in post-pubertal women. The clinical features of NCAH and PCOS overlap. These features include hirsutism, irregular menses, chronic anovulation, infertility, acne, and polycystic ovary morphology. Women with NCAH have normal female external genitalia at birth, but may develop clitoromegaly. Affected men are generally identified through family studies or by infertility due to oligospermia.

Individuals with NCAH can carry mutations typically associated with NCAH on both alleles or a mutation associated with NCAH on one allele and a more severe mutation on the other allele. The frequency of genotype with nonclassic mutations on both alleles has been reported to range from 25% to 50%. Mutations associated with NCAH include V281L, P453S, and R339H. The P30L mutation is often detected in patients with NCAH, but is typically associated with more severe androgen excess.

Diagnosis   An elevated 17-OHP concentration is the hallmark of 21-OHD. For newborns, random 17-OHP concentrations are greater than 5000 ng/dL (150 nmol/L) after 48 hours of life. Since CAH is an autosomal recessive disorder, confirmatory 17-OHP concentrations need to be associated with mutations on both alleles. Beyond infancy, the blood sample for 17-OHP is best obtained in the early morning because of the circadian rhythm of ACTH and cortisol. In menstruating women, the sample should be obtained in the follicular phase. Morning 17-OHP concentrations greater than 200 ng/dL are not diagnostic, but warrant further evaluation.

When random 17-OHP concentrations are equivocal, an ACTH stimulation test should be performed. Following collection of a baseline blood sample, synthetic ACTH, 250 μg is administered. A second blood sample is obtained at 30 or 60 minutes for progesterone, 17-hydroxypregnenolone, 17-OHP, DHEA, androstenedione, and cortisol determinations. Since 17-OHP concentrations may be elevated in other disorders of steroidogenesis such as 11β-hydroxylase deficiency, determinations of deoxycorticosterone and 11-deoxycortisol concentrations may be warranted. In general, ACTH-stimulated 17-OHP concentrations

greater than 1500 ng/dL (45 nmol/L) are associated with *CYP21A2* mutations on both alleles. However, there are reports of individuals with diagnostic genotypes and ACTH-stimulated 17-OHP values between 1000 and 1400 ng/dL (30-45 nmol/L). Individuals with CAH have elevated 21-deoxycortisol concentrations. Indeed, screening by 21-deoxycortisol would improve the diagnostic evaluation for 21-OHD, but commercial availability of this hormone assay is limited. Steroid quantification, by LC-MS/MS has demonstrated elevated 17-OHP, 21-deoxycortisol, 16α-hydroxyprogesterone, and progesterone; these steroids comprise sensitive and specific biomarkers to accurately identify patients with CAH due to 21-OHD.

Mineralocorticoid deficiency is characterized by hyponatremia, hyperkalemia, and elevated plasma renin activity (PRA). In the affected newborn, the low dietary salt intake unmasks the mineralocorticoid deficiency within 7 to 14 days of life. Infants also tend to have a relative decreased sensitivity to aldosterone. Hyperkalemia typically occurs first with subsequent fall in sodium and bicarbonate. Individuals with simple virilizing CAH may have elevated PRA without overt salt loss. Individuals with NCAH generally have normal electrolytes and PRA values.

Additional helpful studies in the infant with ambiguous genitalia include karyotype and pelvic ultrasound to assess for Müllerian-duct derived structures. The apparent non-visualization of ovaries on ultrasound can be misleading because ovaries may be too small to be reliably detected by ultrasound.

Outcomes   The tremendous improvement in treatment means that almost all children with CAH thrive and are transitioned to adult care. To date, however, outcome reports are disappointing because many affected adults are lost to medical follow-up. In general, adults with CAH are shorter than expected for mid-parental height predictions. Both women and men can experience infertility. Increased incidences of obesity, dyslipidemia, osteoporosis, cardiovascular risk factors, and insulin resistance have been reported. Although various aspects of QoL are negatively impacted, some individuals appear to do relatively well. A study from a single medical center reported that adverse health outcomes, with the exception of testicular adrenal rest tumors (TARTs), were largely related to long-term disease control.

Factors contributing to the lack of fecundity include vaginal stenosis with dyspareunia, impaired sensation, poor self-esteem, and disinterest in having children. Reproductive outcome for women with CAH can be greatly improved by adequate suppression of progesterone and 17-OHP to promote ovulation and implantation of the fertilized ovum; this may require optimizing both glucocorticoid and mineralocorticoid therapies. In one series, women with CAH reported later sexual debut, fewer pregnancies and children, and increased incidence of homosexuality, which were related to type of surgical correction and the severity of their mutations. Another series utilizing a questionnaire provides more optimistic data regarding outcome for adult women with virilizing CAH. The consequences of prenatal androgen exposure on the developing female brain are being explored.

In addition to gonadal adrenal rest tumors, the elevated adrenal C19 steroid secretion can suppress gonadotropin secretion

resulting in hypogonadotropic hypogonadism and subsequent oligospermia. Gonadal adrenal rest tumors, predominantly TARTs occur in up to half of men. These tumors arise from adrenal cells that descend with the testes during testicular development. TARTs are not malignant, but can compress the rete testis and seminiferous tubules culminating in testicular atrophy and obstructive azoospermia. Ultrasound and MRI are helpful to detect TARTs less than 2 cm because small lesions are generally not palpable. TARTs may be present in childhood and adolescence and have rarely been described in men with nonclassic CAH. Although TARTs have been attributed to poor adherence, pathogenesis of TARTS may be more complicated. Ovarian adrenal rest tumors (OARTs) have been infrequently reported in affected women.

11β-Hydroxylase Deficiency

CAH due to 11β-hydroxylase deficiency (OMIM 202010) is characterized by glucocorticoid deficiency, excessive androgen secretion, and hypertension. This form of CAH is associated with mutations in the 11β-hydroxylase (*CYP11B1*) gene. The *CYP11B1* gene is located at chromosome 8q22 in close proximity to a highly homologous gene *CYP11B2*, which codes for aldosterone synthase. *CYP11B1* is primarily expressed in the zona fasciculata, whereas *CYP11B2* is expressed primarily in the zona glomerulosa.

This form of CAH accounts for 3% to 8% of cases with incidence of 1 in 100,000 in European populations. The highest incidence is reported among Moroccan Jews. The enzyme converts 11-deoxycortisol to cortisol primarily in the zona fasciculata (see Figure 14–5). Loss of function mutations leads to the accumulation of 11-deoxycortisol (compound S) and deoxycorticosterone (DOC). Low renin hypokalemic hypertension is an inconsistent finding. Individuals with 11β-hydroxylase deficiency present similar clinical features as patients with 21-OHD. Affected females may present with ambiguous genitalia. Similar to 21-OHD, a continuum of clinical manifestations occurs, including milder or nonclassic forms. Curiously, infants may experience transient salt loss, presumably due to limited salt intake, and transient mineralocorticoid resistance in the newborn period or during illnesses, while on hormone replacement therapy.

Typical laboratory findings are elevated compound S and DOC concentrations. Serum concentrations of 17-OHP, androstenedione, and testosterone may be mildly elevated. ACTH stimulation testing, which includes measurement of compound S and DOC, may be helpful to assess for this disorder. PRA concentrations are low or suppressed. Using urinary GC-MS, an increase of tetrahydro-11-deoxycortisol (THS) and the precursor to metabolite ratio {THS to [tetrahydrocortisone + tetrahydrocortisol (THF)+5αTHF)]} can be used as disease-specific diagnostic markers.

3β-Hydroxysteroid Dehydrogenase Deficiency

CAH due to 3β-hydroxysteroid dehydrogenase type 2 deficiency (OMIM 201810) is associated with mutations in the *HSD3B2* gene. This gene is mapped to chromosome 1p13.1 where it is located in close proximity to the highly homologous *HSD3B1* gene. The *HSD3B1* gene is expressed primarily in skin, placenta, prostate, and other peripheral tissues. The *HSD3B2* gene is primarily expressed in the adrenal cortex and gonads. Mutations in *HSD3B2*, but not *HSD3B1*, have been detected in patients with 3β-hydroxysteroid dehydrogenase deficiency.

The NAD$^+$-dependent enzyme 3β-hydroxysteroid dehydrogenase/Δ^5-Δ^4-isomerase is a bifunctional enzyme that catalyzes the conversion of the Δ^5 steroid precursors, pregnenolone, 17-hydroxypregnenolone, and DHEA into the respective Δ^4-ketosteroids, progesterone, 17-OHP, and androstenedione (see Figure 14–5). It is, therefore, essential for biosynthesis of mineralocorticoids, glucocorticoids, and sex steroids. Mutations are associated with virilization of the external genitalia of 46,XX fetuses due to increased DHEA synthesis. Affected 46,XY fetuses may have undervirilization of the external genitalia secondary to testosterone deficiency. Despite decreased testosterone synthesis, affected 46,XY fetuses usually have intact Wolffian duct structures (including vas deferens).

Infants with complete loss of function mutations may present in the neonatal period with acute adrenal insufficiency. Typical presentations for the non-salt-losing forms include premature pubarche and perineal hypospadias in affected XY infants. Confirmatory laboratory findings include elevated pregnenolone, 17-hydroxypregnenolone, and DHEA concentrations with elevated ratios of Δ^5 to Δ^4 steroids. Because enzymatic activity of the type 1 isozyme is unimpaired, elevated 17-OHP and androstenedione concentrations may be found.

Treatment

The goals of treatment for the CAHs include provision of adequate glucocorticoid and mineralocorticoid replacement to achieve normalization of adrenal C19 steroids. Finding the balance between overtreatment and undertreatment is often challenging. Titration of hydrocortisone dosing needs to be individualized based on clinical features, linear growth velocity, skeletal maturation, and steroid hormone concentrations in plasma and urine. The dose of mineralocorticoid should be adjusted so that the electrolytes and blood pressure, as well as the PRA, are in the normal range. An *in vitro* study confirmed the clinical observation that 17-OHP and progesterone have anti-mineralocorticoid effects; these steroid hormones interfere with aldosterone-mediated mineralocorticoid receptor transactivation.

Most of this section refers to 21-OHD because it is the most common form of CAH. Treatment of patients with salt-losing adrenal hyperplasia may be separated into chronic and acute phases. The neonate identified by newborn screening should be started on glucocorticoid and mineralocorticoid replacement therapy once confirmatory blood samples have been obtained. Typical hydrocortisone dosage is 25 to 30 mg/m^2/d in the newborn period that is subsequently weaned to approximately 8 to 15 mg/m^2/d. For infants and children diagnosed with simple virilizing CAH, typical hydrocortisone dosage is also approximately 8 to 15 mg/m^2/d. For most patients, the hydrocortisone dose is divided into three doses administered every 8 hours. Reverse circadian dosing with the highest hydrocortisone dose in the evening may benefit some individuals.

Oral fludrocortisone is utilized for mineralocorticoid replacement at 0.05 to 0.3 mg in the newborn. Fludrocortisone is also weaned to doses of approximately 0.05 to 0.2 mg daily. Some infants require added dietary salt (1-3 g daily) using a salt solution prepared by a pharmacist to avoid dose errors.

Treatment of individuals affected with NCAH is somewhat controversial. When diagnosed in children with advanced skeletal maturation, daily glucocorticoid treatment is generally prescribed to decrease adrenal C19 steroids and slow the rate of bone maturation. Since the major symptoms in adolescent girls include irregular menses, hirsutism, and acne, the use of oral contraceptives may be sufficient to control the symptoms. Most adolescent men are asymptomatic and identified through family screening studies. If the ACTH-stimulated peak cortisol response is less than 500 mmol/L (18 µg/dL), consideration should be given to stress dosing during acute illnesses. Women with NCAH who are actively trying to conceive may benefit from glucocorticoid treatment to promote ovulation, implantation, and decrease risk for miscarriages.

In acute adrenal crises, patients can present with hyponatremia, hyperkalemia, hypovolemia, hypoglycemia, acidosis, hypotension, and shock. Normal saline (10 cc/kg in two consecutive boluses) should be administered to patients. Glucose should be administered if the patient is hypoglycemic. Subsequent replacement of glucose, fluid, and electrolytes is calculated on the basis of deficits and standard maintenance requirements. Hydrocortisone sodium succinate (50 mg/m^2) should be given as a bolus followed by stress dosing.

Patients with physiologic stress such as fever, severe vomiting and diarrhea, severe injury, or elective surgery require additional glucocorticoid treatment. If the patient can take oral hydrocortisone, the usual dose is generally tripled during the acute illness, decreased to double dosage as the patient recovers, and eventually to their usual replacement dose. If unable to tolerate oral medications, intramuscular hydrocortisone sodium succinate should be administered as follows: 25 mg, for infants less than 12 months, 50 mg for children ages 1 to 4 years, and 100 mg for all over age 4 years. Parenteral hydrocortisone, at dose of approximately 30 to 50 mg/m^2/d, should be given every 6 hours if the patient is still unable to tolerate oral medications. In this situation, the patient may benefit from evaluation by a physician and intravenous fluids. All caregivers should be prepared to administer intramuscular hydrocortisone injections in emergencies. All patients should wear medical alert notification badges.

Prenatal dexamethasone therapy of pregnant women who are at risk of carrying a female fetus with atypical genitalia has been utilized to decrease the virilization of the external genitalia of affected female infants. To be effective, this therapy must be initiated at 6 to 7 weeks of gestation. However, human outcome studies have reported that prenatal dexamethasone is associated with orofacial clefts, decreased birth weight, poorer verbal working memory, and poorer self-perception of scholastic and social competence. Using an animal model, prenatal dexamethasone exposure has been shown to alter the neurodevelopmental program of CNS development. At the present time, long-term outcome studies for prenatal dexamethasone treatment are not available. Another option for couples at risk for a child with CAH is preimplantation genetic testing. Non-invasive early prenatal diagnosis using cell-free fetal DNA found in the maternal circulation has been reported.

Girls with genital ambiguity due to CAH may benefit from feminizing genitoplasty. The aims of surgery and the potential risks and benefits need to be thoroughly reviewed with the parents. Most importantly, each patient deserves individualized management. The goals of this surgery include an adequate vaginal introitus, separation of urethra from the vaginal introitus, and minimizing erectile tissue while simultaneously preserving glanular innervation and blood supply.

B. Glucocorticoid receptor gene mutation Primary generalized glucocorticoid resistance is characterized by target tissue insensitivity to glucocorticoids. This disorder can be associated with heterozygous loss of function mutations in the glucocorticoid receptor (*NR3C1*) gene (OMIM 615962). Some mutations act through a dominant negative mechanism to interfere with glucocorticoid signal transduction. Glucocorticoid resistance is characterized by high circulating cortisol concentrations without signs and symptoms of Cushing syndrome. The resistance of the hypothalamic-pituitary axis to suppression leads to increased ACTH secretion with increased secretion of adrenal hormones with salt-retaining and androgenic activities. The normal circadian rhythm is maintained, albeit with a higher set-point for ACTH and cortisol concentrations. The spectrum of clinical manifestations ranges from virilization of female fetuses, premature pubarche, hypertension, hypokalemia, and, in females, hirsutism and infertility. Tests useful to confirm the diagnosis include 24-hour urinary-free cortisol determinations and genetic analysis of the *NR3C1* gene. Treatment involves the judicious use of a synthetic glucocorticoid (eg, dexamethasone) to decrease the excessive ACTH, mineralocorticoid, and adrenal androgen concentrations.

C. Placental aromatase deficiency Aromatase is a cytochrome P450 enzyme that converts C19 androgens to C18 estrogens. The aromatase (*CYP19A1*) gene is located at chromosome 15q21.2. Aromatase is an essential enzyme involved in estrogen biosynthesis in adolescents and adults. Expressed in the human placenta, aromatase converts fetal androgens to estrogens to protect the mother from the potential virilizing effects of fetal androgens.

Placental aromatase deficiency is a rare autosomal recessive disorder due to inactivating *CYP19A1* mutations (OMIM 613546). Decreased aromatase activity impairs conversion of androgens to estrogens resulting in increased androgen concentrations. At birth, affected 46,XX infants are virilized with varying degrees of labioscrotal fusion, clitoromegaly, and perineal scrotal hypospadias (see Figure 14–11). Affected 46,XX individuals generally manifest delayed puberty characterized by minimal or absent breast development, primary amenorrhea, hypergonadotropic hypogonadism, multicystic ovaries, and decreased bone mineral density. As would be anticipated, affected 46,XY infants typically have normal internal and external genital development and present after puberty with tall stature, delayed skeletal maturation, decreased bone

mineral density, and infertility. Phenotypic heterogeneity has been described. Treatment involves appropriate estrogen replacement therapy.

During pregnancies with affected fetuses, maternal testosterone, DHT, and androstenedione concentrations are elevated whereas estradiol, estrone, and estriol concentrations are low. These hormone alterations lead to progressive maternal virilization characterized by hirsutism, clitoral hypertrophy, acne, and frontal balding during pregnancy. In the postpartum period, some clinical features of androgen excess regress, and the elevated androgen concentrations return to normal levels. Nonetheless, aromatase deficiency in the infant can occur in the absence of maternal virilization.

D. P450 oxidoreductase deficiency

The clinical features of P450 oxidoreductase deficiency include genital ambiguity, skeletal anomalies, and impaired steroidogenesis. This disorder is associated with mutations in the cytochrome P450 oxidoreductase (*POR*) gene, which is mapped to chromosome 7q11-12. The protein encoded by *POR* functions as a mandatory electron donor from reduced nicotinamide adenine dinucleotide phosphate to microsomal steroidogenic and hepatic P450 enzymes and plays a major role in glucocorticoid and sex steroid synthesis. This protein interacts with other steroidogenic enzymes including 21-hydroxylase, 17α-hydroxylase/17,20-lyase, and aromatase (see Figure 14–5). The skeletal manifestations have been attributed to disrupted interactions with enzymes involved in sterol, for example, 4α-lanosterol demethylase (CYP51A1) and squalene epoxidase, and retinoic acid metabolism, for example, CYP26 isozymes.

At birth, genital ambiguity can be noted in both male and female infants. Insufficient testosterone synthesis contributes to the undervirilization of male infants. Virilization of female fetuses is attributed to shunting of the excessive 17-OHP to the "backdoor pathway" resulting in increased DHT and other androgens. Phenotypic heterogeneity has been reported with varying severity of genital ambiguity, adrenal insufficiency, and skeletal anomalies. Hypertension associated with increased corticosterone metabolites can occur. Differing from the virilizing CAHs, progressive postnatal virilization does not occur. Affected females may have delayed puberty and ovarian cysts. Typical laboratory findings include elevated basal and ACTH-stimulated 17-OHP, low sex steroids, and normal mineralocorticoid concentrations. Urinary GC/MS steroid profiles indicate decreased 21-hydroxylase and decreased 17α-hydroxylase/17,20-lyase activities. Some affected individuals may benefit from daily glucocorticoid replacement therapy. Others may need glucocorticoid treatment only for stress dosing.

During pregnancy, some mothers developed signs associated with androgen excess such as acne, hirsutism, and clitoromegaly. Prenatal diagnosis can be established using urinary steroid metabolite analysis of maternal urine. One series reported that identification of malformations on prenatal ultrasound was predictive of a more severe phenotype.

The skeletal malformations include craniosynostosis, midface hypoplasia, phalangeal anomalies, and synostosis of large joints. Similar skeletal anomalies occur in the Antley-Bixler syndrome, an autosomal dominant disorder associated with mutations in the fibroblast growth factor receptor 2 (*FGFR2*) gene. The *FGFR2* gene is located at chromosome 10q26.13. However, individuals with Antley-Bixler syndrome due to *FGFR2* mutations have normal steroidogenesis.

E. Maternal sources: Androgens and progestogens

In general, the 46,XX fetus is protected from excessive androgen exposure by the ability of the fetal-placental unit to aromatize androgens to estrogens. Nevertheless, exposure of mothers to testosterone, other androgenic steroids, or some synthetic progestational agents during pregnancy can be associated with virilization of the external genitalia of 46,XX fetuses (see Figure 14–11). Norethindrone, ethisterone, norethynodrel, medroxyprogesterone acetate, and danazol have all been implicated in masculinization of the female fetus. In rare instances, a maternal androgen secreting ovarian or adrenal tumor or a luteoma of pregnancy can be associated with virilization. Placental aromatase protects most 46,XX fetuses in cases of maternal CAH. Another potential etiology includes environmental endocrine disruptors, which are exogenous chemicals capable of interfering with one or more aspects of hormone action. The diagnosis of 46,XX DSD arising from transplacental passage of androgenic steroids is largely a diagnosis of exclusion. Surgical correction of the genitalia, if needed, is the only therapy necessary.

MANAGEMENT OF PATIENTS WITH DSD

General Aspects

The management of patients with DSD is complex, due to clinical heterogeneity and the multiple factors influencing sexual outcomes. Because of this high variability in phenotype and in DSD clinical expression and outcome, all members of the medical team must be comfortable with individualizing their approach to each patient and each family. DSD conditions should be diagnosed as specifically as possible at the time of presentation, as etiology can be an important guiding factor in management. However, contrary to other conditions in medicine, arriving at a specific diagnosis does not necessarily dictate a specific care plan. Hence, the approach is fundamentally different from that utilized for most other diseases encountered in medicine, in which the primary focus is on diagnosis by which therapy is then defined. Additionally, a commitment to family education (based on age, maturity, sophistication, education level, and cultural milieu) is paramount to an optimal outcome, with the goal of full disclosure and repeated psychological assessments. A long-term care plan needs to be formulated in concert with the family.

Patients with DSD presenting with genital ambiguity, prenatal ultrasound confusion, or inconsistency of prenatal karyotype and genital development encountered by the clinician at birth require thoughtful deliberation, regarding male or female assignment and the need for or benefit of early genital surgery. Based on the initial karyotype, the differential diagnosis typically is categorized under the heading of karyotype (see Table 14–1). However, one negative aspect of this type of categorization occurs when a patient or family

mistakenly assumes that the karyotype is a type of cryptic code indicating maleness or femaleness or that it defines sex and/or gender. Instead, the parents should come to understand that while, in the usual situation, a female has a 46,XX karyotype, there are situations in which a genetic disturbance can direct differentiation in another, at times unpredictable, direction. In fact, many autosomal genes are found to be disease-causing in many DSDs and their identification can be a helpful factor for sex assignment since the genes on sex chromosomes do not alone determine reproductive system development (eg, 46,XX males).

Genital ambiguity is not considered a medical emergency. Nonetheless, the coexistence of a life-threatening condition, such as adrenal insufficiency, should be considered to allow immediate therapy if needed. Beyond this initial diagnostic urgency, it is important that parents be provided with timely and increasingly detailed information. They may not have previously been aware that such a condition could occur, and they need to be provided a perspective that their child can be expected to have a good life and that the team will work with them to ensure the best possible care for their child. Full disclosure and education are crucial. Sharing information in a progressive and timely way helps parents understand how such a situation could develop, together with goals to equip them with current information to make decisions regarding care and long-term outcome information, and how to identify biases or stigma so they can be avoided while carefully protecting confidentiality. In spite of recent attempts to inform, there is still strong evidence of the stigmatizing effect associated with a DSD diagnosis, and underscores the need to educate repeatedly the patient and family. Psychological family support and counseling should be provided at the time of diagnosis, followed by periodic psychologic reassessment to identify those in need of on-going counseling.

The initial session with the family involves a review of the embryonic differentiation including the common origin of the clitoris and penis in the genital tubercle, of the labia majora and scrotum from the embryonic labioscrotal folds, and of the labia minora and penile urethra from the urethral folds, as well as the development of the cortex and medulla of the primordial gonad into the ovary and testes. These explanations can provide a basis for the understanding of how too little or too much androgen can result in masculinized female genitalia or insufficient masculinization of male genitalia and how gonads may be partially developed or differentiated in more than one direction. Further, internal reproductive system differentiation should also be explained including the need for adequate testosterone for male development and the role of AMH in suppressing female internal development. This is the basis for understanding how a 46,XY patient with testicular development may have uterine and Fallopian tube development, or why an individual with inadequately masculinized external genitalia likely has insufficient ductal development.

The physical development of the infant should be examined with the parents observing so they will be comfortable seeing their baby's genitals. Viewing an enlarged phallus while being reminded that this may represent an enlarged clitoris or an incompletely formed penis; and partially fused labial-scrotal folds, which may represent partial scrotal fusion or excessive fusion of labia majora;

and partially fused urethral folds which may represent an incompletely formed penile urethra or excessive fusion of labial minora not only informs the parents of the status of their infant, but also provides a conceptual framework to avoid mystique and stigma and allows them to understand how such a situation can develop. As laboratory tests become available, they should be interpreted within the context of the other elements of history, physical examination, hormonal and genetic testing.

Parent/patient support and advocacy groups should be identified objectively, recognizing that non-representative biases may occur. The caveat should be that the multidisciplinary team as well as these groups both desire optimal health care and decisions that result in the best possible broadly defined QoL for the patient. The perception should be that the diagnosis and care of DSD individuals, who have a life-long condition, continues to evolve based on better diagnostic tools, including genetic and visualization techniques, refinements in surgery and ever improving outcome data. It is important to highlight to families that existing outcome reports do not take into account all of the multiple factors that impact QoL. Outcome studies usually do not include a representative patient sampling and may overly represent patients with valid complaints and underrepresent those with the best outcomes. It is important to provide the concept that most DSD conditions are medically complex, and because they interface with human and societal views on sexuality, it is not always possible to predict adult outcomes. Parents, and later the individual, need to understand that they may be angry or upset that such a condition exists but they need to avoid the tendency to blame the caretakers without coming to terms with the fact that this condition occurred. As health-care providers, our goal is that individuals with DSD will become well-adjusted contributing members of society.

While all agree under ideal circumstances that a **multidisciplinary team**, including an endocrinologist, surgeon, geneticist, and psychologist work together with other care providers to facilitate on-going care, most tertiary care centers lack the availability of experienced individuals to staff this type of arrangement. This is particularly true for focused psychological support, with this role often being assumed by another team member. While centers of excellence may improve this deficit, fiscal and time constraints often make this option unfeasible in many situations as well.

At initial assessment, whether at infancy or later, a **careful history** should be obtained including family history of genital ambiguity, unexplained infant death, maternal exposure to hormonal or toxic substances and any evidence of virilization during pregnancy, consanguinity, infertility, amenorrhea, hirsutism, and any other endocrinopathies.

A careful **physical examination** should be performed with specific attention to reproductive system development. Ranges of penis and clitoral size are available. Such information should be considered in context, and a designation of an abnormality is inappropriate simply for a clitoris that exceeds or a penis that does not meet these criteria. There should be careful documentation of genital anatomy to provide the basis to assess changes over time and for future outcome assessment. For example, the enlarged clitoris in a 46,XX infant with CAH may be larger at birth than

after a few weeks of adrenal suppression when androgens have declined considerably. Such regression may indeed be dramatic, and such that clitoral surgery may no longer be considered. The presence and size of a mass(es) in the labial-scrotal or inguinal area need to be documented, and these may represent testes, ovotestes or rarely the cervix of the uterus in an infant with PMDS. When palpable the location of testes should be documented as: scrotal, upper scrotal, within the inguinal canal, lower abdominal or rarely ectopic in the femoral and perineal regions. The configuration of the phallus, location of the meatus, and estimation of amount of corporal tissue should be noted, together with the width, stretched length and degree of chordee (if present). While a palpable midline structure on rectal examination suggests a uterus, this type of examination is unnecessary when a pelvic ultrasound is being done. Genital phenotype can be documented, using the Prader scales for virilization of female genitalia as occurs with CAH, or with the Quigley scales, as used in 46,XY DSD individuals. It is important to realize that the Prader scale, specifically the location of the urethral meatus does not necessarily correlate with the extent of the underlying corpus spongiosum fusion. This can only be determined with surgical exploration; hence, predicting the likelihood of successful surgical repair can be made only during surgery. Anogenital distance (from the center of the anus to the junction between the smooth perineal skin and rugated skin of the scrotum) can be seen as an indicator of testosterone exposure during early fetal life. The normal distance is 3 ± 0.7 cm. For females, an anogenital ratio (distance from anus to posterior fourchette/distance from anus to base of clitoris) greater than 0.5 suggests excessive androgen exposure.

The risk of malignant transformation of gonadal tissues must be considered. Such involves primarily germ cell tumors. Tumor risk is related to gonadal maturation, presence of portion of the Y chromosome, and anatomic location for testicular tissue. DSD patients with gonadal dysgenesis are particularly at risk. Precursor lesions can be identified and stained for particular genetic expression, such as SOX9 and FOXL2. Embryonic stem cell markers, particularly OCT3/4, are useful. The risk for development of germ cell tumors is related to intrinsic risk alleles, disrupted germ cell development, and environmental factors. Gonadal biopsy is necessary to identify germ cells tumors or tumor precursors. The age at which a biopsy should be done should depend upon the age range of tumor development within the particular DSD. This generally begins around pubertal age.

A comprehensive view of the perspectives of clinicians, patients, and their families have identified **three primary ethical principles** applying to management of DSD patients. They are to: (1) foster the well-being of the child and future adult, (2) uphold the rights of children and adolescents to participate in and/or self-determine decisions that affect them, and (3) respect the family and parent-child relationship. The goal of achieving a balance between the rights, needs and interests are outlined in 9 recommendations. These recommendations include: (1) careful weighing of therapeutic decisions, (2) the ultimate rights of the parents, (3) determining a binary sex does not automatically ensure the child's well-being, (4) respect of the parents' value system, (5) interventions should be based on exhaustive diagnostic studies and

the best possible prognosis, (6) explicit reasoning and justification is necessary for any intervention when conclusive scientific evidence is lacking, (7) providing information to the child commensurate with age, (8) documentation of treatment information to provide to the individual as an adult, and (9) periodic review of recent medical scientific findings.

These ethical principles do not support a moratorium on all genital and gonadal surgery during infancy or delaying interventions until the individual is old enough to decide, as some would advocate. There is no evidence-based data upon which to base this type of moratorium. Furthermore, this type of one sizes fits all approach would undermine the rights of many families, whose unique circumstances leads to a desire for genital surgery for their child. In addition, the proposal to delay surgical interventions does not address how it would be determined that an individual had attained such an age to make the decision autonomously and ignores the complete lack of knowledge regarding the effect of living with genitalia that are inconsistent with sex of rearing. Such a policy would introduce a new type of dogma reminiscent to that of the optimal gender approach, creating a form of social experimentation where basic parental rights and responsibilities are restricted at the behest of a highly vocal minority that uses intimidation and legal challenging to supersede basic human bioethical principles.

Diagnostic Workup

As previously said, systematic rules cannot be applied in the management of all DSD patients. The diagnostic workup varies according to the age at presentation (newborn, child, adolescent), and the main complaint (ambiguous genitalia, discordance between fetal karyotype and genital appearance, polymalformations, pubertal delay/primary amenorrhea, infertility), and usually includes a variable array of endocrine laboratory determinations, imaging studies, histological examinations, and genetic tests.

A. Newborns with ambiguous genitalia: Initial steps
Owing to its frequency and life-threatening potential, CAH needs to be excluded in newborns with ambiguous genitalia, especially in the absence of palpable gonads. Neonatal screening programs can detect elevated levels of 17-OHP, and may give a rapid clue in the cases of CAH due to 21-OHD. A karyotype should also be performed.

Other hormone results should be analyzed considering the particular ontogeny of the reproductive axis in newborns: normally, gonadotropins, androgens and AMH are low in the first week of life in both sexes. However, serum AMH and testosterone levels above a certain threshold indicate the presence of testicular tissue (Table 14–3). From the second week, all hormones progressively increase following a sexually dimorphic pattern, where LH increases more than FSH in males—the opposite is observed in females—and AMH and testosterone become clearly higher in males. The existence of ovarian tissue cannot be easily proven due to the lack of specific markers; in fact, estradiol does not show a clear sexual dimorphism. All steroid determinations should be

TABLE 14–3 Serum levels of gonadotropins and gonadal hormones in normal newborns.

Hormone	Day 2	Day 7	Day 14	Day 21	Day 30
Testosterone (ng/dL)					
Boys	20-100	30-120	30-190	30-190	120-350
Girls	20-40				20-40
AMH (pmol/L)					
Boys	200-550	250-600	300-650	400-700	450-950
Girls	ND-10				ND-30
FSH (IU/L)					
Boys	ND-0.50	0.50-3.50	0.50-3.50	0.70-2.00	0.90-2.20
Girls	ND-0.30				1.00-13.00
LH (IU/L)					
Boys	ND-0.50	0.80-7.00	1.20-4.00	1.20-4.50	1.60-4.50
Girls	ND-0.20				0.20-2.00

performed after extraction procedures to avoid assay interference rendering artificially elevated serum levels.

Genitoscopy and/or genitography are usually needed to establish the morphologic characteristics of the internal genitalia and urinary structures. Due to difficulties with adequate resolution, ultrasonography may give false results, unless performed by a professional with a vast experience in the study of the reproductive tract in newborns.

B. Next steps when karyotype is 46,XX (Table 14–4)
Patients with non-palpable gonads   Since 17-OHP may not be unequivocally elevated in some forms of CAH, hormonal investigations must be repeated together with studies of salt balance (Na$^+$, K$^+$) and PRA. Hypoglycemia and failure to thrive reflect glucocorticoid deficiency, which also results in lack of negative feedback regulation of POMC/ACTH; consequently, ACTH is elevated, and genitalia and nipples are hyperpigmented. The diagnosis of CAH also needs the distinction between salt-wasting and simple virilizing forms. The finding of high levels of PRA, together with hyponatremia and hyperkalemia, is indicative of aldosterone insufficiency, which usually presents in the second or third week of life. Salt wasting is not typically evident in 11β-hydroxylase deficiency owing to accumulation of DOC, but it may be triggered by glucocorticoid treatment. Detailed hormonal and genetic studies in patients with CAH have been discussed earlier.

A history of maternal virilization during pregnancy suggests placental aromatase deficiency. No clinical or biochemical signs of adrenal dysfunction are observed in the newborn. In the first 2 weeks of life, estradiol is below the normal female range and androgens and gonadotropins are elevated. Molecular studies should point to the *CYP19A1* gene.

Ovotesticular DSD should be suspected, if there is no history of maternal virilization and levels of androgens and AMH are between the normal female and male ranges (see Table 14–4).

Patients with palpable gonads   Ovotesticular DSD is the most likely diagnosis. Rarely, dysgenetic testicular DSD may

TABLE 14–4 Laboratory findings in 46,XX DSD.

T	AMH	LH/FSH	17OH-P	Cortisol	Na$^+$	K$^+$	Defect
High	Normal	Normal	High	Low	Low	High	CAH (21-OHD, 11β-OHD)
Slightly high	Normal	Normal	Low	Low	Low	High	CAH (3β-HSD)
High/Normal	Normal	High	High	Normal/Low	Normal	Normal	POR deficiency
High	Normal	High	Normal	Normal	Normal	Normal	Aromatase deficiency
High	High	Normal	Normal	Normal	Normal	Normal	Testicular DSD
High	Normal	Normal	High	High	Normal	Low	GR defect
High/Normal	Normal	Normal	Normal	Normal	Normal	Normal	Virilizing tumor (mother)
High	High	High	Normal	Normal	Normal	Normal	Ovotesticular DSD
Normal	Normal	Normal	Normal	Normal	Normal	Normal	Malformative DSD

Note: Serum levels are considered normal, high, or low as compared to reference values in girls.

Abbreviations: 11β-OHD, 11β-hydroxylase deficiency; 17OH-P, 17 hydroxyprogesterone; 21-OHD, 21-hydroxylase deficiency; AMH, anti-Müllerian hormone; CAH, congenital adrenal hyperplasia; FSH, follicle-stimulating hormone; GR, Glucocorticoid receptor; LH, luteinizing hormone; POR, P450 oxidoreductase.

occur. Usually the amount of functional testicular tissue correlates with the levels of testosterone and AMH.

The definitive diagnosis of ovotesticular or dysgenetic testicular DSD relies on histological studies of both gonads.

When virilization is complete and both gonads are scrotal and of normal size—a discordance between a fetal karyotype and the genital phenotype being the cause for referral of such newborns—the diagnosis of testicular DSD is almost certain. In these cases, serum gonadotropins, testosterone, and AMH are within the normal male range.

Molecular studies should include the search for *SRY* sequences, for *SOX9*, *SOX3*, or *SOX10* duplications, or *RSPO1* or *WNT4* mutations in *SRY*-negative cases.

C. Next steps when karyotype is 46,XY (Table 14–5)

The occurrence of undervirilized genitalia in 46,XY newborns

indicates that either the gonads have not produced enough testosterone or the target organs are not responsive. The assessment of serum testosterone and AMH are very useful for an initial classification into three distinct groups: (1) dysgenetic DSD, (2) isolated androgen deficiency, and (3) androgen end-organ defects. Anatomical assessment by imaging studies and surgical procedures show the existence of uterus and Fallopian tubes in the first case, and their absence in the remaining two.

Low testosterone and AMH: testicular dysgenesis Serum levels of testosterone and AMH below the normal male range for age indicate that both Leydig and Sertoli cell populations are affected (ie, dysgenetic DSD). FSH and LH are elevated, reflecting a lack of negative feedback. The amount of functional testicular tissue usually correlates with the degree of virilization. When the newborn is completely feminized and has come to medical

TABLE 14–5 Laboratory findings in 46,XY or chromosomal DSD with a Y lineage.

T	AMH	LH	FSH	Other Hormonal Features	Diagnosis
Undetectable No response to hCG	Undetectable	High	High	Normal adrenal steroids	Complete gonadal dysgenesis
Low Poor response to hCG	Low	High	High	Normal adrenal steroids	Partial testicular dysgenesis Asymmetric gonadal differentiation
Low/normal Poor/normal response to hCG	Low/normal	Norma/High	Normal/High	Normal adrenal steroids	Mild testicular dysgenesis Ovotesticular DSD
Undetectable/low No/poor response to hCG	Normal/High	High	Normal	All gonadal steroids low Normal adrenal steroids	Leydig cell aplasia/hypoplasia
Undetectable/low No/poor response to hCG	Normal/High	High	Normal	All gonadal and adrenal steroids low	Lipoid congenital adrenal hyperplasia (StAR/P450scc)
Low Poor response to hCG	Normal/High	High	Normal	High pregnenolone and progesterone Low cortisol	17α-hydroxylase deficiency POR deficiency
Low Poor response to hCG	Normal/High	High	Normal	High 17OH-pregnenolone and 17OH-progesterone Normal cortisol	17-20 lyase deficiency
Low Poor response to hCG	Normal/High	High	Normal	High pregnenolone, 17OH-pregnenolone and DHEA Low cortisol	3β-HSD type 2 deficiency
Low Poor response to hCG	Normal/High	High	Normal	High Δ4-androstenedione	17β-HSD type 3 deficiency
Normal or High	Normal	Normal	Normal	Low DHT High urinary 5β/5α-reduced metabolite ratios: Tetrahydrocorticosterone/allotetrahydrocorticosterone Etiocholanolone/androsterone	5α-reductase type 2 deficiency
Normal/High	Normal/High	High	Normal	Normal adrenal steroids	AIS
Normal	Normal	Normal	Normal	Normal adrenal steroids	Malformative DSD

Note: Serum levels are considered normal, high or low as compared to reference values in boys.

Abbreviations: 3β-HSD, 3β-hydroxysteroid dehydrogenase; 17β-HSD, 17β-hydroxysteroid dehydrogenase; AMH, anti-Müllerian hormone; AIS, androgen insensitivity syndrome; DHEA, dehydroepiandrosterone; DHT, dihydrotestosterone; hCG, human chorionic gonadotropin; P450c17, Cytochrome P450 17-hydroxylase/17,20-lyase; P450scc, Cytochrome P450 side-chain cleavage; POR, Cytochrome P450 oxidoreductase; StAR, Steroidogenic Acute Regulatory protein.

attention because of discordance with a fetal karyotype, the diagnosis of complete or "pure" gonadal dysgenesis should be considered if AMH is undetectable.

When the genital abnormality is not associated with other malformations, *SRY*, *NR5A1* (coding for SF1), and *MAP3K1* are the most likely candidate genes, together with partial Xp duplications encompassing the DSS region containing *NR0B1* (encoding DAX1). Renal function particularly the occurrence of proteinuria should be explored in these patients. Absence of renal dysfunction excludes *WT1* mutations or deletions.

In patients with associated primary adrenal insufficiency, *SF1* defects should be considered. Patients with campomelic dysplasia are likely to have a mutation in the *SOX9* gene. Polyneuropathy, which might not be easy to detect in a newborn, is suggestive of *DHH* mutations, whereas mental retardation and α-thalassemia should drive the attention to *ATRX*. When mental retardation, craniofacial dysmorphic features and delayed motor development are found, 9p deletions including *DMRT1* should be suspected.

Low testosterone, AMH normal for male: isolated defect of androgen production

When testosterone is low in the first 3 to 6 months of life but AMH is within or above the normal male range, only Leydig function is affected. LH may be elevated and FSH normal. A prolonged hCG test (6 or 7 IM injections of 1500 IU every other day, with gonadal steroids measured in serum 24 to 72 hours after the last injection) is usually needed. When all steroids are low, but adrenal function is normal, the defect is in the LH/hCG receptor. Conversely, if basal ACTH and PRA are elevated, with low Na$^+$ and high K$^+$, StAR or P450scc deficiency is likely. An ACTH test (6 IM injections [0.5 mg/m^2 of body surface area] of depot ACTH every 12 hours, with adrenal steroids measured before the first injection and 12 hours after the last injection) should make the differential diagnosis.

Low testosterone and cortisol responses to ACTH, associated with elevated pregnenolone and progesterone and low 17OH-pregnenolone/pregnenolone and 17OH-progesterone/progesterone ratios, are suggestive of insufficient 17α-hydroxylase activity due to mutation of the gene encoding P450c17 (see Figure 14–5). Elevated DOC and corticosterone, normal Na$^+$ and K$^+$, and hypertension in older infants are almost confirmatory. The associated occurrence of bone malformations typical of Antley-Bixler syndrome is suggestive of P450 oxidoreductase (POR) deficiency. An isolated deficiency of the 17,20-desmolase activity of P450c17 is suspected when adrenal function is normal, testosterone is low, and the ratios DHEA/17OH-pregnenolone and Δ4-androstenedione/17OH-progesterone are reduced.

More rarely, testosterone and cortisol may show a low response, in association with decreased stimulated ratios of Δ4/Δ5 steroids (progesterone/pregnenolone, 17OH-progesterone/17OH-pregnenolone and Δ4-androstenedione/DHEA). Here, 3β-HSD type 2 deficiency is suspected (see Figure 14–5), even if absolute values of Δ4 steroids (progesterone, 17OH-progesterone, and Δ4-androstenedione) are normal or somewhat elevated.

When testosterone is low but all precursors are elevated, resulting in a low testosterone/Δ4-androstenedione ratio, 17β-HSD type 3 deficiency is most likely (see Figure 14–5).

Male-range testosterone and AMH: androgen end-organ defects

Fetal-onset hypogonadism can be ruled out when both testosterone and AMH are within or above the male range. Androgen target-organ defects, including androgen insensitivity and deficient DHT production, should be suspected (see Figure 14–9).

CAIS is most likely in newborns with completely female external genitalia. Conversely, in patients with ambiguous genitalia, the differential diagnosis between PAIS and 5α-reductase type 2 deficiency is not straightforward. A high testosterone/DHT ratio after an hCG test is suggestive of 5α-reductase type 2 deficiency; however, false negatives may result due to DHT conversion by 5α-reductase type 1 activity, particularly in the infant. On the other hand, false positive results may result from reduced DHT levels observed in PAIS patients; this is due to the fact that 5α-reductase activity is regulated by androgens via the androgen receptor. Urinary 5β/5α-reduced metabolite ratios have a better diagnostic efficiency: high tetrahydrocorticosterone/allotetrahydrocorticosterone and high etiocholanolone/androsterone ratios confirm 5α-reductase type 2 deficiency. Molecular genetic testing may be helpful in this situation.

D. Next steps when karyotype is atypical (Table 14–5)

Sex chromosome DSD, resulting in gonadal dysgenesis containing variable amounts of testicular tissue, is the diagnosis in a newborn with ambiguous genitalia and a sex chromosome mosaicism with at least one cell line carrying a Y chromosome. In patients 45,X/46,XY, 45,X/46,XY/47,XXY, or other similar mosaicisms, gonadal dysgenesis—either symmetrical or asymmetrical (mixed gonadal dysgenesis)—is likely. In patients with 46,XX/46,XY mosaicisms, ovotesticular DSD is the rule. The comments made above for those disorders (see karyotypes 46,XX and 46,XY) apply also here.

E. 46,XX or 46,XY newborns with polymalformative syndromes (see Tables 14–4 and 14–5)

The occurrence of consistent features of undervirilization (ie, micropenis, hypospadias, bifid scrotum, cryptorchidism) in a 46,XY patient suggests that testicular dysgenesis is part of the malformation syndrome. Low testosterone and AMH, together with elevated gonadotropins, confirm the diagnosis. Conversely, when inconsistency is noted between the different genital structures (eg, isolated penile agenesis with normal scrotum and palpable testes), gonadal dysfunction is unlikely. Normal testosterone, AMH, and gonadotropins are found. In some patients, where the malformations of the abdominopelvic wall or the anogenital region interfere with the anatomic assessment of the genitalia, testosterone and AMH determinations allow the establishment of the gonadal sex.

F. Genetic testing in DSD

Families and health-care personnel are interested to establish the genetic and molecular basis in the patient with a disorder of sex development. With the development of newer and high-throughput technologies, it is increasingly possible to identify the specific genetic etiology in an individual patient. Availability of this information decreases anxiety for families and improves the ability of health-care providers to predict

the phenotype of the patient and genetic recurrence risks. Genetic testing is increasingly used in medical decision-making. Although the individual patient's management may not change, identification of a causative variant may predict recurrence risk and can be used for preimplantation genetic diagnosis. In many instances, genetic evaluation of the parents, affected relatives, and other unaffected relatives is essential to ascertain if a mutation is *de novo* and if it segregates with the patient's phenotype. Variants are categorized by likely functional significance and consequences (Table 14–6). With the vast amount of genetic data being collected, public databases have been developed to provide information regarding clinical phenotype, functional significance, and prevalence. One caveat is that the information listed in these databases (Table 14–7) may not have been updated or recently reorganized.

Peripheral blood karyotype analyses remain useful to detect balanced chromosomal translocations and large chromosomal anomalies such as structural rearrangements of the X chromosome. Fluorescent in-situ hybridization (FISH) enhances routine karyotype analyses because it can characterize unknown marker chromosomes and chromosomal rearrangements. For the newborn with ambiguous genitalia, rapid turn-around times to detect X and Y chromosomes may be used to help direct subsequent studies.

Comparative genomic hybridization (CGH) arrays are a high-resolution method that can detect submicroscopic genetic abnormalities. This method can identify copy number variants (CNVs) and detect microdeletions and microduplications. In some instances, these findings represent *de novo* submicroscopic chromosomal rearrangements. However, low level mosaicism and balanced chromosomal rearrangements such as translocation or inversions may overlooked by this technique. CGH has become

useful to identify novel candidate genes associated with DSD. Alterations in gene dosage have been associated with anomalies in gonadal development. The use of customized CGH targeted for CNV associated with DSD can interrogate multiple genes simultaneously which expedites the diagnostic process. During childhood, the phenotypes of patients with mutations in *AR*, *LHCGR*, and *CYP17A1* can overlap. Targeted CGH may be able to differentiate among these genetic disorders and eliminate the need to perform hCG stimulation tests.

TABLE 14–6 Terms used to characterize sequence variants.

Term	Meaning
Pathogenic	Contributes to disease
Likely pathogenic/ implicated	Data consistent with pathogenic role with a defined level of confidence
Benign	Available data indicates that variant does not contribute to disease
Likely benign	Data consistent with non-pathogenic role with a defined level of confidence
Variant of uncertain significance (VUS)	Data inadequate to classify variant
Associated	Variant is significantly enriched in affected individuals compared to matched controls
Damaging	Variant affects the normal levels or biochemical function of a gene product
Deleterious	Variant reduces reproductive fitness of carriers

TABLE 14–7 Useful databases for genetics.

Database	URL	Features
1000 Genomes Project	http://browser.1000genomes.org	Database of variants from 26 populations
OMIM	www.omim.org	Database of human genes and genetic disorders started by Dr. Victor McKusick in 1960s
Human Gene Mutation Database	www.hgmd.org	Database of annotated literature
DECIPHER	http://decipher.sanger.ac.uk	Molecular cytogenic database linking genomic microarray data with phenotype
NCBI Genome	http://www.ncbi.nlm.nih.gov/genome	Full human genome reference sequences
NCBI-dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP	Repository for variants and genomic position. This is a collection of variants
Gene Cards	http://www.genecards.org/	Searchable database for annotated human genes
USCS Genome Bioinformatics	http://genome.ucsc.edu/index.html	Searchable database for annotated human genes
HuGE Navigator	http://www.cdc.gov/genomics/hugenet/hugenavigator.htm	Continuously updated knowledge base in human genome epidemiology supported by the Centers for Disease Control and Prevention
GeneReviews	http://www.ncbi.nlm.nih.gov/books/NBK1116/	Peer-reviewed descriptions of inherited diseases in standardized format
ClinGene	https://www.clinicalgenome.org/	Partnership between public, academic, and private institutions to establish genetic resources for improved patient care

Genome-wide association studies (GWAS) enable a hypothesis-free approach to identifying genes associated with specific disorders. With this method, allele frequencies of affected persons are compared to controls (case-control analyses). All individuals in both groups are genotyped for a large number of common single-nucleotide polymorphisms (SNPs). The goal is to identify sequence variants associated with the specific phenotype of interest. Many of the associated variants are located outside of the coding region of a gene complicating interpretation of the functional significance. These SNPs may affect gene regulatory elements, influence local chromatin structure, or affect cofactor recruitment. Non-coding genes including microRNAs and long noncoding RNAs need to be considered. Replication of findings is essential, especially for GWAS, to minimize false positive associations due to differences in population architecture.

Whole exome sequencing (WES) is a technique in which the coding regions (exons) of thousands of genes are analyzed simultaneously using next-generation sequencing techniques. Multiple variations in one individual's DNA sequence can be detected by comparing the patient's exome to a normal reference sequence. Technically, WES is accomplished by breaking the DNA into small fragments, sequencing the fragments, aligning the sequences, and using computer software to compare the aligned sequences to the human genome reference sequence. This technique can detect single nucleotide changes and small insertions and deletions. Results are screened for rare variants, variants previously associated with disease, and variants predicted to generate a functional consequence. Whole genome sequencing (WGS) targets the entire genome. Both WES and WGS are hypothesis-free approaches that generate much data with the need for powerful bioinformatics tools. WES provides more coverage of exons and, currently, is less costly than WGS. Typically, WES detects 15,000-20,000 variants and WGS detects 3 to 4 million variants that differ from the human reference genome. However, WES will not identify variants involving non-coding regions. WGS has greater sensitivity to detect small structural variations (eg, insertions, deletions, and translocation). Neither technique adequately identifies larger CNVs, repetitive sequences such as trinucleotide repeats, aneuploidy, or epigenetic alterations. For both techniques, detected genetic alterations need to be validated by polymerase-chain reaction amplification followed by Sanger sequencing or restriction fragment length polymorphism analysis.

Ethics of WES and WGS   These powerful tools can identify genetic variants known to be associated with diseases unrelated to the disorder of interest. Therefore, pretest counseling and informed consent are essential components of the process involved in performing genetic studies. Parents and patients need to be advised that incidental findings unrelated to the reason for testing may be detected. Many, but not all, variants reported to be associated with genetic disorders represent robust associations. However, false assignments of causality remain a major challenge. Clinician and researchers are faced with the responsibility, obligation, and pressure of assessing the functional significance of genetic variants with the potential consequence of misdiagnosis.

Sex Assignment

With the primary goal of sex assignment being an assignment that aligns well with the future gender identity, it is crucial to realize that future gender development cannot be predicted with complete certainty. At best, a probable gender identity is all that can be developed. This uncertainty of gender development, as is increasingly publicly chronicled in non-DSD transgendered persons, creates a formidable barrier to success. Factors to be considered are based on: etiologic diagnoses (when known); degree of external genital development; surgical options; anticipated quality of sexual function (based upon the two previous factors); fertility potential; malignancy risk; familial/social/cultural/psychosocial factors; psychosocial risks for the individual and parents; and degree of fetal androgen exposure. It is important to realize that familial/social/cultural/psychosocial factors; psychosocial risks for the individual and parents, and degree of fetal androgen exposure may on occasion override the gender development anticipated for a specific diagnosis. It is clear that there is a need to limit aggressive surgical removal of potentially fertile gonadal tissue or potentially sexually responsive phallic tissue during infancy.

The general principle for sex assignment is, when the differentiation is primarily female, the assignment should be female, and when male, male. Nonetheless, at all levels of sex assignment, a certain degree of uncertainty is inescapable. This should be shared with parents to instill the flexibility needed to raise a child with DSD with understanding and compassion. However, in each patient, there must be careful consideration to determine whether there is a reason to consider an exception. Major exceptions include when differentiation was not primarily toward one or the other sex, as exemplified by some, but not all, cases of ovotesticular DSD or when there is diminished androgen responsiveness as in some with PAIS. Phenotype variability among those with the same enzyme deficiency and/or diagnosis can be considerable, such that sex assignment must consider all principles on which assignment is based and should not rely on an etiologic diagnosis alone.

The principles used for assignment should carry the greatest probability of adult happiness and success. The concepts include: (1) alignment of physical sex with gender identity; (2) capacity for satisfying sexual function and response; (3) potential for "normal" genitalia; and (4) fertility but only when there is a reasonable probability of alignment of sex assignment with adult gender identity. Since the probabilities may vary within an individual for these key considerations, it is important to consider the composite evidence in a compassionate, flexible and unique way that offers the highest likelihood of adult happiness and success.

The impact of fetal androgen exposure for a female fetus must be judged with care when considering sex assignment with the caveat that those 46,XX infants with a virilized genital tubercle have been exposed to male typical levels of androgens in utero. Evidence of the magnitude of effect is indirect since the impact upon the developing CNS appears to occur mid-gestation. It is important to understand that the extent of external genitalia masculinization reflects androgen levels during the first trimester and that circulating androgen levels obtained after birth reflect greater

hypothalamic-pituitary-gonadal activity during early infancy and that neither is a good surrogate measure of mid-gestational androgen exposure. Furthermore, the impact of androgen levels on sexual identity is not robust beyond the more male type gender role behavior (rough and tumble childhood behavior) seen in children exposed to high levels of androgen in utero. A relationship between degree of masculinization and male type behavior is assumed. As evidenced among most 46,XX DSD patients with classical CAH, the impact of fetal androgen exposure upon gender identity development or sexual orientation is certainly not direct or linear.

The viewpoint needs to shift from the concept of karyotype as definitive evidence for sex assignment to the realization that mutations on genes on sex chromosomes and autosomes may override this. Even though DSD nomenclature categorizes groups based on karyotype as a guide to specific etiologies, the presence of XX or XY should not be viewed, particularly by parents, as substantiation for a female or male assignment.

Assignment is not necessarily the same for an etiologic diagnosis since hormone secretion and differentiation varies. The impact of varying levels of fetal androgen exposure to the CNS, expression of pertinent hormone-responsive genes at various developmental windows during fetal life and later during childhood/adolescence, and gene expression that occurs independently of hormone milieu all make accurate projection of gender development problematic.

Evidence for the impact of prenatal androgens indicates that infants are not born gender neutral. With this realization, there has been a general shift toward male assignment for 46,XY individuals with evidence of testicular derived androgen exposure in utero and postnatal androgen responsiveness, even in those with "inadequate" penis development. The shift in assignment among 46,XY patients preceded publication of the DSD consensus statement in 2006, documented by evidence of 35% of such patients being assigned male before 1990 with 68% since 1999. The same reasoning also challenges one to consider a male assignment in 46,XX patients with CAH with fully male external genitalia based upon the favorable outcome data in those who, owing to a delayed diagnosis, were raised male.

A. Sex assignment for 46,XX DSD individuals with genital ambiguity

The long-term perspective has been that patients with 46,XX DSD, the majority of whom have 21-hydroxylase deficient CAH, should always have a female assignment. This reasoning was based upon the following considerations. (1) Although the external genitalia may be sufficiently virilized to require surgery for functionality and cosmesis, the outcome, when genital sensitivity is preserved, is expected to be satisfactory. (2) Since internal reproductive structures, (ovaries, Fallopian tubes and uterus) are differentiated normally, fertility would be expected after proper treatment. (3) While these patients are exposed to sufficiently elevated fetal androgen levels to result in a tendency to male gender role behaviors, particularly during childhood, gender identity is usually female.

However, it is appropriate to include a consideration of a male assignment among 46,XX patients with essentially male genitalia,

including those with CAH and sex reversal with significant virilization. Outcome data have been obtained from patients who, owing to the presence of near normal male genitalia at birth, were considered male and subsequently identified with the XX karyotype only well after the newborn period. These individuals, though infertile, manifest a clear male gender identity and a reasonable QoL. Between 1896 and 2002, 46 cases assigned male at birth without reassignment were reported. Subsequently 12 46,XX CAH men were reported with male gender identity, sexual orientation toward females, normal male libido, masculinity, frequent satisfactory (orgasmic) sexual activity, and satisfactory body image. Currently, there are many other 46,XX CAH patients being reared with a male sex assignment, as a consequence of parents' choice.

The 2006 consensus statement indicated that all 46,XX CAH patients should be assigned female based on available outcome data showing that 95% identify as female. In failing to address the 5% who do not identify as females, an outcome that has been shown to be related to degree of masculinization of the genitalia, the consensus statement has promoted an algorithmic approach to DSD management. In contrast, we feel a male assignment should be considered for 46,XX, patients with CAH, as in other DSD conditions, in which external genitalia are essentially male and there is evidence of high levels of in utero androgen exposure, while taking into account the social and cultural environment.

B. Sex assignment for 46,XY DSD individuals with genital ambiguity

As noted earlier, among 46,XY DSD patients, there has been a nearly universal shift to male assignment rather than female assignment, when there is indirect evidence of CNS androgen exposure during fetal life and demonstration of androgen responsiveness. This shift was heavily influenced by the male gender development among those 46,XY patients with cloacal exstrophy who had been assigned female at birth and were then gonadectomized. Additionally adult males with very small penises almost invariably identify as male.

Outcome information, particularly QoL, for 46,XY DSD conditions have been inadequately studied. Female assignment is certainly appropriate for those with female genitalia (no genital ambiguity) including CAIS, pure gonadal dysgenesis and LH receptor defect. Because a significant portion of 46,XY patients with DSD are given a diagnosis of PAIS, male assignment should occur only after demonstration of androgen responsiveness.

C. Specific considerations

1. Female assignment remains appropriate for those with CAIS, but the clinician must realize that psychological distress issues exist and need to be recognized and addressed.

2. The recommended sex assignment recommended for those with 5α-reductase deficiency is male, based upon the report that the majority develop a male gender identity when gonads are left intact. However, it would be appropriate to individualize the approach since fertility may be a problem for those raised male. Those raised female have indicated satisfactory sexual function. Categorical male sex assignment for 5α-reductase or 17β-HSD deficiencies, each of which shows

heterogeneous development, carries the risk of repeating the previous generations' mistake of too rigidly applying accepted concepts and should be challenged.

3. Male assignment is also recommended for those with 17β-HSD deficiency, with testes left intact, based upon the development of male gender identity among the majority. For those with this diagnosis, it is appropriate to question whether the recommendation for all should be based upon the outcomes for only a portion of this population without consideration for the variability in the severity of the phenotype. Both those raised males and female have indicated satisfactory sexual function, while those raised male have indicated dissatisfaction with penis length and those raised female have been reported to have clinical distress.

D. Sex assignment consideration for those with variable karyotype, gonadal, external and internal differentiation

Sex assignment for those having ovotesticular DSD and mixed gonadal dysgenesis (or asymmetric gonadal differentiation) continue to require complete assessment regarding the extent of masculine or feminine development, even unilateral, as male or female, considering the basic consideration of potential for fertility, gonadal hormonal secretion, and genital development potentially enabling sexual function.

For ovotesticular DSD, if ovarian tissue containing follicles is present together with a unilateral Fallopian tube, a uterus and some vaginal development, a female assignment may be considered; while if phallic development is judged sufficient to become a functioning penis and testicular tissue containing Leydig cells and germ cells is present, a male assignment could be considered. An important consideration is to retain functional tissue when possible in the event that the patient, whose gender outcome is typically uncertain, wishes to self-reassign gender later in life. When considering fetal CNS androgen exposure, it is crucial to remember that while the extent of masculinization of the external genitalia is evidence of androgen secretion at some phases of fetal life, it does not necessarily correlate with androgen exposure and hence masculinizing effects in the CNS.

Long-term Outcomes

The goal of the multifaceted therapy for patients with DSDs is to minimize the consequences and enhance the ability to live healthy, contented lives and be productive persons with fulfilling social and personal relationships. While care currently is focused upon outcome, guidelines for use in childhood/adolescence can, at best, be generalized because medical, surgical, and psychological care is continually changing and because current outcome data reflect treatment given about two decades ago. Hence, outcome studies from that time are not completely applicable because care has changed based on evolving new information. Perceptions of outcome have also been heavily influenced by intersex activist groups that seek to generalize anecdotal outcomes and often inappropriately provoke negativity because of highly publicized cases with poor outcome in the absence of adequate evidence-based data. Furthermore, the implication promoted by some activist groups that any genital surgery is "torture" and comparable to primitive female circumcision is unfairly provocative. If taken to the groups' intended conclusion, this would impose a belief system onto families/patients for whom surgery may benefit their unique circumstance.

Existing outcome studies, involving endocrine outcomes, tumor development, psychological, surgical, fertility, urinary, and sexual outcomes, do not take into account the multiple factors impacting outcome, nor do they account for changes in the lives of the study subjects that may enhance or hinder satisfaction. An individual's outcomes are related to personality characteristics including the ability to adjust to/accept their situation, and the availability of social support particularly within their families. Outcome reports often involve small numbers of patients, particularly within arbitrary diagnostic categories that are not physiologically or psychologically related, without appropriate control groups, lacking verified tools and without analysis of multiple impacting facets.

However, in the absence of highly quality outcome data, inferences from available studies are pertinent. A summary of outcomes studies published since 2012 include multiple diagnoses (Table 14–8). It is important to realize that there will never be outcome studies that will result in clear directives for managing individuals in most areas of sexual development, because the underlying conditions persist for life. The ultimate QoL depends upon ability to accept that this is a lifelong condition associated with many factors. Better tools are being developed to assess outcome, and multidisciplinary groups have been formed, but the need to individualize our care will remain as the key guideline for the development of a care plan that seeks to maximize the chance for a "good" outcome.

Satisfactory clinical outcomes are more difficult for those with more severe developmental anomalies. Those with the greatest genital ambiguity and insufficiency have the greatest risk for poor outcome. Hence, outcome data should be assessed based upon percentages of those with a given diagnosis considering acceptance of genital appearance cosmetically; perception of quality of sexual function; and availability, quality and consistency of intimate, familial and general social support. Factors need to be individualized and applied for each individual, including the initial phenotype of external and internal genitalia, physical health, social and psychological adjustment, mental QoL, social participation and support, coping skills, treatment satisfaction following outcome of reconstructive surgery, reproductive system function, and overall QoL. In this regard, it is noteworthy that half of parents expressed regret regarding their decision to have distal hypospadias repaired in their sons long before they were physically mature. Variations in perceptions of learning their DSD diagnosis by young women include lack of memory of details concerning their diagnosis but with positive recollections of the clinician's flexibility, adaptability, and support. Later concerns involved potential social, practical, and philosophical issues.

Because of the variation within diagnoses, different factors beyond being raised as male or female impact QoL. It is recognized that a crucial, albeit incompletely understood factor, is fetal androgen exposure to the CNS. This varies among etiologic diagnoses such 5α-reductase and P450c17 deficiency. The interpretation of a fetal androgen exposure concept is complicated because no good index of fetal CNS androgen exposure currently exists.

TABLE 14–8 Summary of outcome reports from 2012.

46,XX DSD

Most with severe ambiguity had surgery

Extent of virilization and number of surgeries affected cosmetic outcome

Sexual function unsatisfactory because of lack of experience and sexuality-related distress

For vaginal hypoplasia, many had short vaginal length, sexual dysfunction, sexual distress

Among CAH:

    Surgical outcome positive regarding appearance but less so for sexual function

    Level of confluence of vagina into urogenital sinus, a major detriment for cosmetic outcome with more surgeries

    For those sexually active, lower functioning re: desire, arousal, pain, but not satisfaction or orgasm, greater sexual distress compared with reference group

    Evidence of good outcome: Self-satisfaction, Happiness indices, Body image perception

    Sex life satisfaction related to family support and medical care

46,XY DSD

Male sex preferable

Among those raised female, sex life dissatisfaction among 1/3 to 1/2 regarding desire, arousal, dyspareunia, and surgical outcome

Among those raised male:

    Genital appearance and function dissatisfaction jeopardizing sexual Quality of Life (QoL)

    Persistence of small penis size with a portion with length insufficient to achieve penetration

    Psychological support-integral part of management

    Body image and psychosexual not different than controls

Among parents of boys with distal hypospadias: regret regarding decision for surgery associated with postoperative complications

Among CAIS after sigmoid colon vaginaplasty: good subjective outcome scores, with rare stenosis and reoperative rates, satisfaction among sexually active

Ovotesticular DSD

Male: female assignment 2:1

Surgery and hormonal replacement need variable based on anatomic development

Significant male urethral post-surgical complications

Sexual activity more frequent among males

Combined diagnoses

Seventy-five percent satisfied with sex assignment, remainder includes those with uncertainty, transgender issues, and identification with neither male or female

Neovagina satisfactory among those with CAIS and vaginal agenesis, 80% married or sexually active

Reduced health-related QoL: self-esteem, physical wellbeing, school functioning, not associated with atypical gender role behavior, but associated with gender dysphoria

In Indonesian population, most without medical or surgical treatment, gender change common, especially female to male among masculinized 46,XY individuals, cross-dressing behavior among girls and dissatisfaction among adolescent/adult females

---

While there may be predictable outcomes related to the underlying condition, other outcomes and comorbidities with medical care and social inputs cannot be predicted. For example, the impact of transition from the pediatric setting to the less holistic approach is uncertain.

Studies of psychological outcomes of DSD remain problematic for several key reasons: lack of standardized or systematic approach, small sample size, heterogeneous populations that are not physiologically related, and the results that extend across a broad range of patient types. Since assessment involves self and sexual identity and image, perception of genital development, and peer and romantic relationships, studies may also be hampered by subject's sense of shame and invasion of intimacy, embarrassment, reticence, perceived stigma, lack of assurance of confidentiality, cultural constraints, and the particular point in time of assessment. The importance of family support and acceptance, the individual's understanding of his/her condition and good quality medical and surgical care are correlated with positive self-satisfaction and body image perception, happiness indices, and satisfactory sexual lives. Overall, the most must be made of what is available while remaining cognizant of study limitations and the risk of applying general principles to individual cases and the resultant loss of an individualized approach.

Recent reports have found the following: In one case series satisfactory gender identity was found among 75%, with 24% identifying with both genders and 3% with neither. Nine percent changed gender and 14% had evidence of increased transgender scores; an endpoint of health-related QoL (HRQOL) among 4 DSD subgroups found no differences in QoL for 46,XX and 46,XY DSD individuals raised male or female, and found no association between HRQOL and atypical gender role behaviors. However, this study did show that CAH females (46,XX) with CAH and gender dysphoria had lower HRQOL scores and that the percent of patients with cross-gender parental responses or cross-gender behaviors was higher than in control girls; in contrast, CAH boys (46,XY) were no different from control boys in the response rates for cross-gender parental responses or cross gender behaviors. Cross-gender behavior was exhibited in 12.5% of the girls with CAH. These data suggest that female patients with CAH in general do well unless they manifest gender dysphoria. Does it also suggest a generalized higher risk for poor outcomes among those exposed to high levels of androgens during fetal life who are assigned females?

Overall the sexual QoL for individuals with both 46,XX and 46,XY DSD until recently has been reported as impaired. Dissatisfaction involves non-sexuality, intercourse frequency, avoidance, anorgasmia, and penetration difficulties, all of which may relate to the underlying atypical development, medical/surgical care or both, but this dissatisfaction also extends to the ability or opportunity to develop social, including romantic and intimate, relationships. Among 46,XX DSD females, overall sexual function remains unsatisfactory, and there is a strong association between severity of ambiguity and the level of the confluence of the urethra and the vagina and the favorability of the cosmetic outcome. Genital surgery did seem to improve appearance, but did not seem to improve sexual functioning among women with CAH more than 15 years of age. Obstacles to a good adult outcome were noted in those with vaginal hypoplasia, in which short vaginal length often persisted after therapy and was associated with sexual dysfunction.

Those with 46,XY DSD conditions cited arousal and orgasm problems, sexual anxiety, communication problems, less sexual activity, and dissatisfaction with function and appearance. While male assignment is now common for most 46,XY patients, some

DSD males often have a penis that is insufficient for penetrative intercourse, and this remains strongly associated with negative sexual outcomes and sexual dissatisfaction in this group. While 46,XY DSD males with micropenis prefer the male gender, they cite diminished penis size as having a negative impact on sexual QoL. While body image indices did not differ from controls, satisfaction with genital appearance is different among men with 46,XY DSDs. Penis length was compromised, sexual function poor, and a significant portion with severe hypospadias were unable to achieve vaginal penetration. Sexual outcomes in three groups of 46,XY DSD patients found significant dissatisfaction among some: those non-virilized (including CAIS) cited lack of desire/arousal and dyspareunia, those virilized but raised female complained of poor surgical outcome and difficulties with clitoral arousal, while the virilized group raised male complained of poor genital function. While male assignment seems to avoid the high risk of gender dysphoria in virilized 46,XY DSD patients, the outcomes remain less than optimal because of surgical/medical limitations on increasing penis size and the strong social/cultural pressure of penetrative intercourse as a measure for male sexual fitness.

Outcome varies within diagnostic categories related to severity of genital ambiguity and whether anatomic development is adequate to enhance a positive body image and sexual function. Hence, a male gender is currently frequently chosen in spite of concerns regarding penis development. The issue of diminished penis size persists. Most with 46,XY DSD report satisfaction with long-term results of masculinizing genitoplasty, but complaints relate to sexual activity and urinary symptoms. While most are satisfied with their sex assignment, the majority indicates that they feel uncomfortable with the opposite sex: issues like penis size, problems with erections and ejaculation, voice pitch and facial hair vary considerably.

The goal for surgical outcome has generally been good cosmetic appearance and potential for sexual experience with sufficient sensitivity for satisfactory responsiveness. Most data are from the relatively common DSD, 46,XX CAH. In this group, fertility is diminished and is correlated with extent of masculinization of external genitalia at birth. Functional outcomes of vaginoplasty, female gender identity, and attitudes regarding self and life in general are reported to be satisfactory among the majority in this group. Women with CAH who had nerve-sparing genital surgery, rather than traditional surgical approaches, had greater clitoral sensitivity. The ability of achieve orgasm was not different from controls nor was this orgasmic ability correlated with clitoral sensitivity. Satisfaction with sexual life was not different from controls, except for those with complete loss of function mutations on both alleles. In the latter group, satisfaction with sexual life was not correlated with ability to orgasm. Satisfaction, including sexual function, is reported after vaginoplasty with sigmoid colon and jejunal flap. After surgery for distal, mid, and proximal hypospadias, men had more urinary problems, shorter penile length and less satisfaction with appearance compared with comparably aged control men. These differences were greater among those having had proximal hypospadias. Among those with ovotesticular DSD, a portion of both males and females were sexually active, while

those assigned male had significant more post-surgical urethral complications.

An evaluation of Indonesian children, adolescents, and adults with DSDs, most of whom had not had medical or surgical treatment reports, findings that may be related to inadequate assessment and care but also to cultural factors, including the role of men and women in society. Gender change, primarily from female to male, primarily involving those with 46,XY karyotype and virilization was frequent among adults (44%). Gender identity behavior problems and cross-dressing behavior was more frequent among girls than boys.

In summary, reports indicate that anatomic insufficiencies, even after current therapies, may not be completely corrected. DSD conditions are lifelong with persistent consequences. While good family and peer relationships are supportive, outcome data indicate a persistence of sociocultural issues. Evidence suggests better cosmetic and functional genitalia after surgery and that satisfaction with sexual function is not primarily related to genital functionality including sensitivity. Hence, psychology factors including self-acceptance and good body image are at least as important as medical and surgical care. Finally, while outcome data for diagnostic groups provide general guidelines, the need for an individualized approach to promote a good outcome must be emphasized.

Fertility Issues

Fertility potential in DSD patients should be analyzed considering the clinical form and sex assignment. Here we will address fertility issues in patients with 46,XY DSD, 46,XX DSD (excluding CAH, dealt with earlier) and sex chromosome DSD (excluding Klinefelter and Turner syndromes, which have been described in detail in other chapters).

A. 46,XY DSD   Patients with complete forms of 46,XY DSD are raised as females. In those cases resulting from **pure gonadal dysgenesis**, although the uterus and Fallopian tubes are present, spontaneous fertility is not possible due to the lack of oocytes and ovarian steroids. Pregnancy might be achieved after replacement therapy and oocyte donation.

In **severe defects of androgen synthesis** or in **CAIS**, Müllerian derivatives are absent. Nonetheless, the successful pregnancy recently reported in a patient with congenital absence of the uterus (Rokitansky syndrome) following uterine transplantation has opened a promising alternative.

As already discussed, partial forms of 46,XY DSD result in a broad phenotypic spectrum. In patients raised as males, fertility might be disturbed for several reasons. Spermatogenic defects are primarily present in patients with **partial testicular dysgenesis**, but may also be seen with **partial defects of androgen synthesis** and **PAIS**. Pubertal development among these patients may be partial, requiring testosterone replacement, and there is an increased risk of germ cell tumors. Longstanding cryptorchidism, anatomic defects—like hypospadias or malformations of the epididymis or vas deferens—and complications of genitourinary surgery are other factors affecting fertility. Most patients

are infertile, but there are reports of fatherhood in patients with testicular dysgenesis due to mutations in *NR5A1*, encoding SF1, and in *MAP3K1*, or with steroidogenic defects.

DHT synthesis defects represent a particular case because DHT has a major role in the masculinization of the urogenital sinus and external genitalia but not in spermatogenesis. Patients with 5α-reductase type 2 deficiency virilize spontaneously at puberty due to 5α-reductase type 1 activity. Adult patients usually have male libido and sexual activity if normal intercourse is not impaired by small penis size. Although semen analysis is characterized by low volume due to poor function of the prostate and seminal vesicles, sperm production occurs due to normal intratesticular testosterone concentration, and fertility has been reported.

In patients with mutations in *AMH* or its receptor *AMHR2*, azoospermia is frequent due to the longstanding cryptorchidism or to damage of testicular blood supply during surgical procedures.

B. 46,XX DSD   As mentioned earlier, 46,XX DSD due to CAH raises particular considerations regarding fertility, which are discussed in detail in the section of this chapter on CAH.

Patients with **aromatase deficiency** are born with ambiguous genitalia and progress to hypergonadotropic hypogonadism with pubertal delay, primary amenorrhea, and hyperandrogenism. Enlarged ovaries with cysts, prone to torsion, have been reported. These factors, along with surgical consequences of genital reconstruction, might affect reproduction but longterm consequences on fertility are unknown.

In patients with **ovotesticular DSD**, the testicular tissue is usually more or less dysgenetic. If the patient is reared as male, the considerations are similar to those previously described for 46,XY partial dysgenesis. The ovarian tissue usually develops more normally. Therefore, in patients raised as females the preservation of ovarian tissue is of major importance. Pregnancy has been reported, either spontaneously or with assisted reproductive techniques, in a limited number of cases. Unfortunately, many of them have had complications, like preterm labor or morbidity related to the delivery process, due to the female reproductive tract defects.

Patients with 46,XX testicular DSD usually have normal testicular function until mid-puberty with spontaneous progression of secondary sex characteristics. However, the existence of two X chromosomes and the absence of a Y chromosome drive to massive germ cell degeneration, small testes and azoospermia.

C. Sex chromosome DSD   Patients with sex chromosome aneuploidies and/or mosaicisms usually have dysgenetic gonads, and the considerations already made for other types of dysgenetic DSD apply. Recently, one case of paternity has been reported after successful intracytoplasmic sperm injection from a 46,XX/46,XY patient raised as male.

REFERENCES

Normal Sex Differentiation

Bergadá I, Milani C, Bedecarrás P, et al. Time course of the serum gonadotropin surge, inhibins, and anti-Mullerian hormone in normal newborn males during the first month of life. *J Clin Endocrinol Metab*. 2006;91:4092-4098.

Biason-Lauber A, Chaboissier MC. Ovarian development and disease: the known and the unexpected. *Semin Cell Dev Biol*. 2015;45:59-67.

Chemes HE, Rey RA, Nistal M, et al. Physiological androgen insensitivity of the fetal, neonatal, and early infantile testis is explained by the ontogeny of the androgen receptor expression in Sertoli cells. *J Clin Endocrinol Metab*. 2008;93:4408-4412.

Flück CE, Pandey AV. Steroidogenesis of the testis—new genes and pathways. *Ann Endocrinol (Paris)*. 2014;75:40-47.

Jost A. Problems of fetal endocrinology: the gonadal and hypophyseal hormones. *Recent Prog Horm Res*. 1953;8:379-418.

Klonisch T, Fowler PA, Hombach-Klonisch S. Molecular and genetic regulation of testis descent and external genitalia development. *Dev Biol*. 2004; 270:1-18.

Lasala C, Carré-Eusèbe D, Picard JY, Rey R. Subcellular and molecular mechanisms regulating anti-Mullerian hormone gene expression in mammalian and nonmammalian species. *DNA Cell Biol*. 2004;23:572-585.

Lin YT, Capel B. Cell fate commitment during mammalian sex determination. *Curr Opin Genet Dev*. 2015;32:144-152.

Rey RA, Josso N, Racine C. 2016 Sexual differentiation. In: De Groot LJ, ed. *Endotext*. S. Dartmouth, MA, USA: MDText.com, Inc.; Updated 12 June 2016. Accessed at: http://www.endotext.org/pediatrics/pediatrics2017/pediatricsframe2017.htm, 1 March, 2017.

Salazar-Martinez E, Romano-Riquer P, Yanez-Marquez E, Longnecker MP, Hernandez-Avila M. Anogenital distance in human male and female newborns: a descriptive, cross-sectional study. *Environ Health*. 2004;3:8.

Sekido R, Lovell-Badge R. Genetic control of testis development. *Sex Dev*. 2013;7:21-32.

Turcu A, Smith JM, Auchus R, Rainey WE. Adrenal androgens and androgen precursors-definition, synthesis, regulation and physiologic actions. *Compr Physiol*. 2014;4:1369-1381.

Ungewitter EK, Yao HH. How to make a gonad: cellular mechanisms governing formation of the testes and ovaries. *Sex Dev*. 2013;7:7-20.

Classification and General Pathogenesis of DSD

Achermann JC, Domenice S, Bachega TA, Nishi MY, Mendonça BB. Disorders of sex development: effect of molecular diagnostics. *Nat Rev Endocrinol*. 2015;11:478-488.

Ahmed SF, Hughes IA. The genetics of male undermasculinization. *Clin Endocrinol (Oxf)*. 2002;56:1-18.

Auchus RJ, Chang AY. 46,XX DSD: the masculinised female. *Best Pract Res Clin Endocrinol Metab*. 2010;24:219-242.

Houk CP, Lee PA. Intersexed states: diagnosis and management. *Endocrinol Metab Clin North Am*. 2005;34:791-810, xi.

Lee PA, Houk CP, Ahmed SF, Hughes IA. In collaboration with the participants in the International Consensus Conference on Intersex organized by the Lawson Wilkins Pediatric Endocrine Society and the European Society for Paediatric Endocrinology. Consensus statement on management of intersex disorders. *Pediatrics*. 2006;118:e488-500.

Rey RA, Grinspon RP. Normal male sexual differentiation and aetiology of disorders of sex development. *Best Pract Res Clin Endocrinol Metab*. 2011;25:221-238.

Dysgenetic DSD

Barbaro M, Balsamo A, Anderlid BM, et al. Characterization of deletions at 9p affecting the candidate regions for sex reversal and deletion 9p syndrome by MLPA. *Eur J Hum Genet*. 2009;17:1439-1447.

Bondy CA, Group TSS. Care of girls and women with Turner syndrome: a guideline of the Turner Syndrome Study Group. *J Clin Endocrinol Metab*. 2007;92:10-25.

Conte FA, Grumbach MM, Kaplan SL. A diphasic pattern of gonadotropin secretion in patients with the syndrome of gonadal dysgenesis. *J Clin Endocrinol Metab*. 1975;40:670-674.

El-Khairi R, Achermann JC. Steroidogenic factor-1 and human disease. *Semin Reprod Med*. 2012;30:374-381.

Groth KA, Skakkebæk NE, Host C, Gravholt CH, Bojesen A. Klinefelter syndrome—a clinical update. *J Clin Endocrinol Metab*. 2013;98:20-30.

Loke J, Pearlman A, Radi O, et al. Mutations in MAP3K1 tilt the balance from SOX9/FGF9 to WNT/beta-catenin signaling. *Hum Mol Genet*. 2014;23:1073-1083.

Ratcliffe S. Long term outcome in children of sex chromosome abnormalities. *Arch Dis Child*. 1999;80:192-195.

Reynaud K, Cortvrindt R, Verlinde F, De Schepper J, Bourgain C, Smitz J. Number of ovarian follicles in human fetuses with the 45,X karyotype. *Fertil Steril*. 2004;81:1112-1119.

Sharpe RM, Skakkebæk NE. Testicular dysgenesis syndrome: mechanistic insights and potential new downstream effects. *Fertil Steril*. 2008;89:e33-38.

Disorders of Steroidogenesis

Auchus RJ, Miller WL. Defects in androgen biosynthesis causing 46,XY disorders of sexual development. *Semin Reprod Med*. 2012;30:417-426.

Costa EM, Domenice S, Sircili MH, Inacio M, Mendonça BB. DSD due to 5alpha-reductase 2 deficiency—from diagnosis to long term outcome. *Semin Reprod Med*. 2012;30:427-431.

Flück CE, Meyer-Boni M, Pandey AV, et al. Why boys will be boys: two pathways of fetal testicular androgen biosynthesis are needed for male sexual differentiation. *Am J Hum Genet*. 2011;89:201-218.

Flück CE, Tajima T, Pandey AV, et al. Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. *Nat Genet*. 2004;36:228-230.

Latronico AC, Arnhold IJ. Inactivating mutations of the human luteinizing hormone receptor in both sexes. *Semin Reprod Med*. 2012;30:382-386.

Mendonça BB, Costa EM, Belgorosky A, Rivarola MA, Domenice S. 46,XY DSD due to impaired androgen production. *Best Pract Res Clin Endocrinol Metab*. 2010;24:243-262.

Miller WL. P450 oxidoreductase deficiency: a disorder of steroidogenesis with multiple clinical manifestations. *Sci Signal*. 2012;5:pt11.

Taylor AE, Keevil B, Huhtaniemi IT. Mass spectrometry and immunoassay: how to measure steroid hormones today and tomorrow. *Eur J Endocrinol*. 2015;173:D1-12.

Congenital Adrenal Hyperplasia

Barbaro M, Soardi FC, Ostberg LJ, et al. In vitro functional studies of rare CYP21A2 mutations and establishment of an activity gradient for nonclassic mutations improve phenotype predictions in congenital adrenal hyperplasia. *Clin Endocrinol (Oxf)*. 2015;82:37-44.

Bidet M, Bellanne-Chantelot C, Galand-Portier MB, et al. Clinical and molecular characterization of a cohort of 161 unrelated women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency and 330 family members. *J Clin Endocrinol Metab*. 2009;94:1570-1578.

Finkielstain GP, Kim MS, Sinaii N, et al. Clinical characteristics of a cohort of 244 patients with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2012;97:4429-4438.

Forest MG. Age-related response of plasma testosterone, delta 4-androstenedione, and cortisol to adrenocorticotropin in infants, children, and adults. *J Clin Endocrinol Metab*. 1978;47:931-937.

Kamrath C, Hartmann MF, Wudy SA. The balance of cortisol-cortisone interconversion is shifted towards cortisol in neonates with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Steroid Biochem Mol Biol*. 2014;143:386-391.

Khattab A, Yuen T, Sun L, et al. Noninvasive prenatal diagnosis of congenital adrenal hyperplasia. *Endocr Dev*. 2016;30:37-41.

Krone N, Braun A, Roscher AA, Knorr D, Schwarz HP. Predicting phenotype in steroid 21-hydroxylase deficiency? Comprehensive genotyping in 155 unrelated, well defined patients from southern Germany. *J Clin Endocrinol Metab*. 2000;85:1059-1065.

Krone N, Reisch N, Idkowiak J, et al. Genotype-phenotype analysis in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *J Clin Endocrinol Metab*. 2012;97:E257-267.

Livadas S, Dracopoulou M, Dastamani A, et al. The spectrum of clinical, hormonal and molecular findings in 280 individuals with nonclassical congenital adrenal hyperplasia caused by mutations of the CYP21A2 gene. *Clin Endocrinol (Oxf)*. 2015;82:543-549.

Miller WL, Witchel SF. Prenatal treatment of congenital adrenal hyperplasia: risks outweigh benefits. *Am J Obstet Gynecol*. 2013;208:354-359.

Moran C, Azziz R, Carmina E, et al. 21-Hydroxylase-deficient nonclassic adrenal hyperplasia is a progressive disorder: a multicenter study. *Am J Obstet Gynecol*. 2000;183:1468-1474.

New MI, Carlson A, Obeid J, et al. Prenatal diagnosis for congenital adrenal hyperplasia in 532 pregnancies. *J Clin Endocrinol Metab*. 2001;86:5651-5657.

Reisch N, Hogler W, Parajes S, et al. A diagnosis not to be missed: nonclassic steroid 11beta-hydroxylase deficiency presenting with premature adrenarche and hirsutism. *J Clin Endocrinol Metab*. 2013;98:E1620-1625.

Speiser PW. A Rationale for Mineralocorticoid Supplementation in Classic Congenital Adrenal Hyperplasia. *Horm Res Paediatr*. 2015;83:422-423.

Speiser PW, Azziz R, Baskin LS, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2010;95:4133-4160.

Tardy-Guidollet V, Menassa R, Costa JM, et al. New management strategy of pregnancies at risk of congenital adrenal hyperplasia using fetal sex determination in maternal serum: French cohort of 258 cases (2002-2011). *J Clin Endocrinol Metab*. 2014;99:1180-1188.

Turcu AF, Rege J, Chomic R, et al. Profiles of 21-Carbon Steroids in 21-hydroxylase Deficiency. *J Clin Endocrinol Metab*. 2015;100:2283-2290.

Other Forms of DSD

Belgorosky A, Guercio G, Pepe C, Saraco N, Rivarola MA. Genetic and clinical spectrum of aromatase deficiency in infancy, childhood and adolescence. *Horm Res*. 2009;72:321-330.

Biason-Lauber A, Konrad D, Navratil F, Schoenle EJ. A WNT4 mutation associated with Mullerian-duct regression and virilization in a 46,XX woman. *N Engl J Med*. 2004;351:792-798.

Grinspon RP, Rey RA. When hormone defects cannot explain it: malformative disorders of sex development. *Birth Defects Res C Embryo Today*. 2014;102:359-373.

Grinspon RP, Rey RA. Disorders of sex development with testicular differentiation in SRY-negative 46,XX individuals: clinical and genetic aspects. *Sex Dev*. 2016;10:1-11.

Hughes IA, Davies JD, Bunch TI, Pasterski V, Mastroyannopoulou K, MacDougall J. Androgen insensitivity syndrome. *Lancet*. 2012;380:1419-1428.

Josso N, Rey R, Picard JY. Testicular anti-Mullerian hormone: clinical applications in DSD. *Semin Reprod Med*. 2012;30:364-373.

Nicolaides NC, Charmandari E. Chrousos syndrome: from molecular pathogenesis to therapeutic management. *Eur J Clin Invest*. 2015;45:504-514.

Vilain E. The genetics of ovotesticular disorders of sex development. *Adv Exp Med Biol*. 2011;707:105-106.

Sex Assignment, Long-term Outcomes and Quality of Life

Bachelot A, Golmard JL, Dulon J, et al. Determining clinical and biological indicators for health outcomes in adult patients with childhood onset of congenital adrenal hyperplasia. *Eur J Endocrinol*. 2015;173:175-184.

Braga LH, Pippi Salle JL. Congenital adrenal hyperplasia: a critical appraisal of the evolution of feminizing genitoplasty and the controversies surrounding gender reassignment. *Eur J Pediatr Surg*. 2009;19:203-210.

Callens N, De Cuypere G, Van Hoecke E, et al. Sexual quality of life after hormonal and surgical treatment, including phalloplasty, in men with micropenis: a review. *J Sex Med*. 2013;10:2890-2903.

Callens N, De Cuypere G, Wolffenbuttel KP, et al. Long-term psychosexual and anatomical outcome after vaginal dilation or vaginoplasty: a comparative study. *J Sex Med*. 2012;9:1842-1851.

Casteras A, De SP, Rumsby G, Conway GS. Reassessing fecundity in women with classical congenital adrenal hyperplasia (CAH): normal pregnancy rate but reduced fertility rate. *Clin Endocrinol (Oxf)*. 2009;70:833-837.

Cools M, Looijenga LH, Wolffenbuttel KP, T'Sjoen G. Managing the risk of germ cell tumourigenesis in disorders of sex development patients. *Endocr Dev*. 2014;27:185-196.

Ediati A, Juniarto AZ, Birnie E, Drop SL, Faradz SM, Dessens AB. Body image and sexuality in Indonesian adults with a disorder of sex development (DSD). *J Sex Res*. 2015;52:15-29.

Guercio G, Costanzo M, Grinspon RP, Rey RA. Fertility Issues in Disorders of Sex Development. *Endocrinol Metab Clin North Am*. 2015;44:867-881.

Han TS, Walker BR, Arlt W, Ross RJ. Treatment and health outcomes in adults with congenital adrenal hyperplasia. *Nat Rev Endocrinol*. 2014;10:115-124.

Houk CP, Lee PA. Approach to assigning gender in 46,XX congenital adrenal hyperplasia with male external genitalia: replacing dogmatism with pragmatism. *J Clin Endocrinol Metab*. 2010;95:4501-4508.

Jurgensen M, Lux A, Wien SB, Kleinemeier E, Hiort O, Thyen U. Health-related quality of life in children with disorders of sex development (DSD). *Eur J Pediatr*. 2014;173:893-903.

Kanhere M, Fuqua J, Rink R, Houk C, Mauger D, Lee PA. Psychosexual development and quality of life outcomes in females with congenital adrenal hyperplasia. *Int J Pediatr Endocrinol*. 2015;2015:21.

Kohler B, Kleinemeier E, Lux A, Hiort O, Gruters A, Thyen U. Satisfaction with genital surgery and sexual life of adults with XY disorders of sex development: results from the German clinical evaluation study. *J Clin Endocrinol Metab*. 2012;97:577-588.

Kolesinska Z, Ahmed SF, Niedziela M, et al. Changes over time in sex assignment for disorders of sex development. *Pediatrics*. 2014;134:e710-715.

Lee PA, Nordenström A, Houk CP, et al. Global Update: Disorders of Sex Development Update since 2006: Perceptions, Approach and Care. *Horm Res Paediatr*. 2016;85:158-180.

Lundberg T, Roen K, Hirschberg AL, Frisen L. "It's Part of Me, Not All of Me": Young Women's Experiences of Receiving a Diagnosis Related to Diverse Sex Development. *J Pediatr Adolesc Gynecol*. 2016;29(4):338-343.

Moran C, Azziz R, Weintrob N, et al. Reproductive outcome of women with 21-hydroxylase-deficient nonclassic adrenal hyperplasia. *J Clin Endocrinol Metab*. 2006;91:3451-3456.

Nordenskjold A, Holmdahl G, Frisen L, et al. Type of mutation and surgical procedure affect long-term quality of life for women with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2008;93:380-386.

Nordenstrom A, Frisen L, Falhammar H, et al. Sexual function and surgical outcome in women with congenital adrenal hyperplasia due to CYP21A2 deficiency: clinical perspective and the patients' perception. *J Clin Endocrinol Metab*. 2010;95:3633-3640.

Pasterski V, Zucker KJ, Hindmarsh PC, et al. Increased Cross-Gender Identification Independent of Gender Role Behavior in Girls with Congenital Adrenal Hyperplasia: Results from a Standardized Assessment of 4- to 11-Year-Old Children. *Arch Sex Behav*. 2015;44:1363-1375.

Reisch N, Rottenkolber M, Greifenstein A, et al. Testicular adrenal rest tumors develop independently of long-term disease control: a longitudinal analysis of 50 adult men with congenital adrenal hyperplasia due to classic 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2013;98:E1820-1826.

Rolston AM, Gardner M, Vilain E, Sandberg DE. Parental Reports of Stigma Associated with Child's Disorder of Sex Development. *Int J Endocrinol*. 2015;2015:980121.

Sanches SA, Wiegers TA, Otten BJ, Claahsen-van der Grinten HL. Physical, social and societal functioning of children with congenital adrenal hyperplasia (CAH) and their parents, in a Dutch population. *Int J Pediatr Endocrinol*. 2012;2012:2.

Schonbucher V, Schweizer K, Rustige L, Schutzmann K, Brunner F, Richter-Appelt H. Sexual quality of life of individuals with 46,XY disorders of sex development. *J Sex Med*. 2012;9:3154-3170.

Strandqvist A, Falhammar H, Lichtenstein P, et al. Suboptimal psychosocial outcomes in patients with congenital adrenal hyperplasia: epidemiological studies in a nonbiased national cohort in Sweden. *J Clin Endocrinol Metab*. 2014;99:1425-1432.

van der Zwan YG, Janssen EH, Callens N, et al. Severity of virilization is associated with cosmetic appearance and sexual function in women with congenital adrenal hyperplasia: a cross-sectional study. *J Sex Med*. 2013;10:866-875.

Wiesemann C, Ude-Koeller S, Sinnecker GH, Thyen U. Ethical principles and recommendations for the medical management of differences of sex development (DSD)/intersex in children and adolescents. *Eur J Pediatr*. 2010; 169:671-679.

Wisniewski AB, Mazur T. 46,XY DSD with Female or Ambiguous External Genitalia at Birth due to Androgen Insensitivity Syndrome, 5alpha-Reductase-2 Deficiency, or 17beta-Hydroxysteroid Dehydrogenase Deficiency: A Review of Quality of Life Outcomes. *Int J Pediatr Endocrinol*. 2009;2009:567430.

Puberty

Dennis Styne, MD

ACTH	Adrenocorticotropin	**IGF-1**	Insulin-like growth factor 1
AZF	Azoospermia factor	**KAL1**	Kallmann syndrome
cAMP	Cyclic adenosine monophosphate	**LDL**	Low-density lipoprotein
DHEA	Dehydroepiandrosterone	**LH**	Luteinizing hormone
DHEAS	Dehydroepiandrosterone sulfate	*LHX3*	*LIM homeobox gene 3*
FSH	Follicle-stimulating hormone	*MRKN3*	*Makorin ring finger protein 3*
GABA	Gamma amino butyric acid	**OMIM**	Online Mendelian Inheritance in Man
GH	Growth hormone	**PCOS**	Polycystic ovary syndrome
GnRH	Gonadotropin-releasing hormone	*PROP-1*	*Prophet of PIT 1*
hCG	Human chorionic gonadotropin	**PROK2**	Prokineticin receptor 2
HDL	High-density lipoprotein	**PRL**	Prolactin
HESX-1	Hesx-1 homeodomain	**PSA**	Prostate-specific antigen
hGH	Human growth hormone	*SF-1*	*Steroidogenic factor 1*
HPLC	High performance liquid	**SHBG**	Sex hormone–binding globulin
MS/MS	chromatography tandem mass spectroscopy	*SHOX*	*Short stature homeobox*
		SRIF	Somatostatin
ICMA	Immunochemiluminometric assay	**TGF-α**	Transforming growth factor alpha

Puberty is best considered as one stage in the continuing process of growth and development that begins during gestation and continues until the end of reproductive life. After an interval of childhood quiescence—the juvenile pause—the hypothalamic pulse generator increases activity in the peripubertal period, just before the physical changes of puberty commence. This leads to increased secretion of pituitary gonadotropins and, subsequently, gonadal sex steroids that bring about secondary sexual development, the pubertal growth spurt, and fertility. Historical records show that the age at menarche in girls in Western countries has steadily declined over the last several hundred years; this is probably due to improvements in socioeconomic conditions, nutrition, and, therefore, the general state of health during that period. This trend appears to be continuing but now due to the effects of the obesity epidemic rather than the improvement in general health.

Many endogenous and exogenous factors can alter age at onset of puberty. While obesity may decrease the age of onset of puberty in girls, chronic illness and malnutrition often delay puberty. There is a significant concordance of age at menarche between mother-daughter pairs and within ethnic populations, indicating the influence of genetic factors. Recent study of genetic loci associated with the age of onset of puberty identified several genes that are likely involved in the regulation of menarche and puberty.

PHYSIOLOGY OF PUBERTY

Physical Changes Associated with Puberty

Descriptive standards proposed by Tanner for assessing pubertal development in males and females are in wide use (denoted as Sexual Maturation stages or, often, Tanner stages). They focus attention on specific details of the examination and make it possible to objectively record subtle progression of secondary sexual development that may otherwise be overlooked. Self-assessment of pubertal development by subjects using reference pictures is used in clinical studies but reliability is less than that achieved by physical examination.

A. Female changes   An early sign of puberty in the female, as noted in longitudinal studies, is an increase in height velocity that heralds the beginning of the pubertal growth spurt; girls are not usually examined frequently enough to demonstrate this change in clinical practice, so breast development is the first sign of puberty noted by most examiners. Breast development (Figure 15–1) is stimulated chiefly by ovarian estrogen secretion, although other hormones also play a part. The size and shape of the breasts may
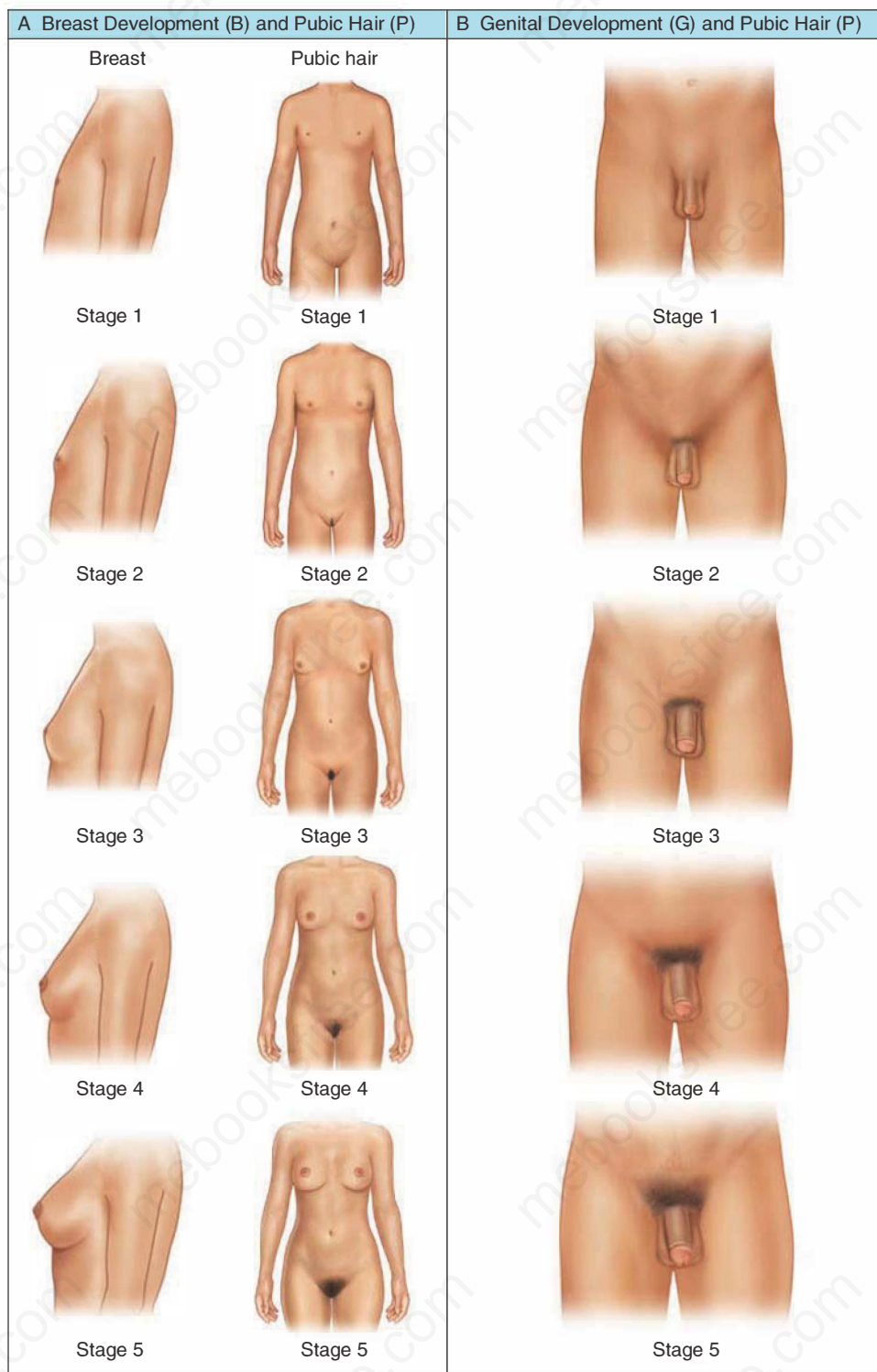
FIGURE 15–1 **(A)** Stages of breast development (B) and female pubic hair development (P), according to Marshall and Tanner. *Stage B1*: Preadolescent; elevation of papilla only. *Stage B2*: Breast bud stage; elevation of breast and papilla as a small mound, and enlargement of areolar diameter. *Stage B3*: Further enlargement of breast and areola with no separation of their contours. *Stage B4*: Projection of areola and papilla

be determined by genetic and nutritional factors, but the characteristics of the stages in Figure 15–1A are similar in all females. Standards are available for the change in areolar (nipple) plateau diameter during puberty: nipple diameter changes little from stages B1 to B3 (mean of 3-4 mm) but enlarges substantially in subsequent stages (mean of 7.4 mm at stage B4 to 10 mm at stage B5), presumably as a result of increased estrogen secretion at the time of menarche. Areolae become more pigmented and erectile as development progresses. Other features reflecting estrogen action include enlargement of the labia minora and majora, dulling of the vaginal mucosa from its prepubertal reddish hue to pink (due to cornification of the vaginal epithelium), and production of a clear or slightly whitish vaginal secretion prior to menarche. Pubic hair development (see Figure 15–1A) is determined chiefly by adrenal and ovarian androgen secretion. Breast development and growth of pubic hair usually proceed at similar rates, but because discrepancies in rates of advancement are possible, it is best to stage breast development separately from pubic hair progression.

Uterine size and shape change with pubertal development as reflected by ultrasonographic studies. With prolonged estrogen stimulation, the fundus-cervix ratio increases, leading to a bulbous form, and the uterus elongates. An endometrial stripe appears with the onset of puberty that is not found in premature thelarche. Ovaries enlarge with pubertal progression. Small cysts are normally present in prepubertal girls and a multicystic appearance develops during puberty, but the polycystic appearance seen in abnormalities of puberty is not present. Experienced ultrasonographers can determine the developmental stage of the uterus and ovaries by comparing measurements with established standards.

It is important to note that girls achieve reproductive maturity prior to physical maturity and certainly prior to psychologic maturity.

B. Male changes The first sign of normal puberty in boys is an increase in the size of the testes to more than 2.5 cm in the longest diameter, excluding the epididymis: this is equivalent to a testicular volume of 4 mL or more. Most of the increase in testicular size is due to seminiferous tubular development secondary to stimulation by follicle-stimulating hormone (FSH), with a smaller component due to Leydig cell stimulation by luteinizing hormone (LH). Thus, if only Leydig cells are stimulated, as with a human chorionic gonadotropin (hCG)-secreting tumor, the testis does not grow as large as in normal puberty. Pubic hair development is brought about by adrenal and testicular androgen secretion and is classified separately from genital development, as noted in Figure 15–1B. A longitudinal study of more than 500 boys suggests adding a stage 2a (absence of pubic hair in the presence of a testicular volume of 3 mL or more) to the classic five stages of pubertal development. Further pubertal development occurred in 82% of the subjects in stage 2a after the passage of 6 months: thus, reaching stage 2a would allow the examiner to reassure a patient that further spontaneous development is likely soon. The appearance of spermatozoa in early morning urinary specimens (spermarche) occurs at a mean chronologic age of 13.4 years or a similar mean bone age; this usually occurs at gonadal stage 3 to 4 and pubic hair stage 2 to 4. Remarkably, spermaturia is more common earlier in puberty than later and boys are reported with spermaturia and no secondary sexual development, suggesting that sperm are directly released into the urine early in puberty while ejaculation may be required for the presence of sperm in the urine of older children.

It is important to note that, like girls, boys achieve reproductive maturity prior to physical maturity and also prior to psychologic maturity.

C. Age at onset Ideally, the upper and lower boundaries encompassing the age at onset of puberty should be set at 2.5 standard deviations (SDs) above and below the mean (encompassing 98.8% of the normal population). Previously, there was no comprehensive study of the start of secondary sexual development adequate to determine the lower limits of normal in U.S. children,

---

to form a secondary mound above the level of the breast. *Stage B5*: Mature stage; projection of papilla only, due to recession of the areola to the general contour of the breast. *Stage P1*: Preadolescent; the vellus over the area is no further developed than that over the anterior abdominal wall (ie, no pubic hair). *Stage P2*: Sparse growth of long, slightly pigmented, downy hair, straight or only slightly curled, appearing chiefly along the labia. This stage is difficult to see on photographs and is subtle. *Stage P3*: Hair is considerably darker, coarser, and curlier. The hair spreads sparsely over the superior junction of the labia majora. *Stage P4*: Hair is now adult in type, but the area covered by it is still considerably smaller than in most adults. There is no spread to the medial surface of the thighs. *Stage P5*: Hair is adult in quantity and type, distributed as an inverse triangle of the classic feminine pattern. Spread is to the medial surface of the thighs but not up the linea alba or elsewhere above the base of the inverse triangle. **(B)** Stages of male genital development (G) and pubic hair development (P), according to Marshall and Tanner. Genital: *Stage G1*: Preadolescent. Testes, scrotum, and penis are about the same size and proportion as in early childhood. *Stage G2*: The scrotum and testes have enlarged, and there is a change in the texture and some reddening of the scrotal skin. There is no enlargement of the penis. *Stage G3*: Growth of the penis has occurred, at first mainly in length but with some increase in breadth; further growth of testes and scrotum. *Stage G4*: Penis further enlarged in length and girth with development of glans. Testes and scrotum further enlarged. The scrotal skin has further darkened. *Stage G5*: Genitalia adult in size and shape. No further enlargement takes place after stage G5 is reached. Pubic hair: *Stage P1*: Preadolescent. The vellus is no further developed than that over the abdominal wall (ie, no pubic hair). *Stage P2*: Sparse growth of long, slightly pigmented, downy hair, straight or only slightly curled, appearing chiefly at the base of the penis. This is subtle. *Stage P3*: Hair is considerably darker, coarser, and curlier and spreads sparsely. *Stage P4*: Hair is now adult in type, but the area it covers is still considerably smaller than in most adults. There is no spread to the medial surface of the thighs. *Stage P5*: Hair is adult in quantity and type, distributed as an inverse triangle. Spread is to the medial surface of the thighs but not up the linea alba or elsewhere above the base of the inverse triangle. Most men have further spread of pubic hair. (Reproduced with permission from Styne DM. *Pediatric Endocrinology, A Clinical Handbook*. Switzerland: Springer International Publishing; 2016.)

as national studies began reporting when children were already 12 years of age or older. Thus, European standards, primarily those of Tanner, were modified for use in the United States. However, a study conducted in medical offices by specially trained pediatricians studying 17,070 girls brought in for routine visits provided information on girls in the United States as young as 3 years. The study revealed that 3% of Caucasian girls reach stage 2 breast development by 6 years of age and 5% by 7 years, whereas 6.4% of African-American girls had stage 2 breast development by 6 years and 15.4% by 7 years. Although this was not a randomly chosen population sampled by longitudinal study (and this led to controversy), it is the largest study available. These data led to the Lawson Wilkins Pediatric Endocrine Society statement that the diagnosis of precocious puberty could be defined as secondary sexual development starting prior to 6 years in African-American girls and prior to 7 years in Caucasian girls who are otherwise healthy. Recent analysis of data from NHANES III demonstrated that children with normal BMI values rarely have breast or pubic hair development before 8 years, although increased BMI in a subset of girls of Hispanic and African American ethnicity led to earlier development. *It is essential to use any such guidelines of the age of onset of puberty only in healthy girls with absolutely no signs of neurologic or other disease that might pathologically advance puberty or a grave diagnosis may be missed.*

Recent data indicate that boys who are overweight have earlier onset of puberty while obese boys have a later onset. However, there has been no change in the standards of the age at onset of puberty of boys in the general population, so 9 years is still taken as the lower limit of normal pubertal development in males, whereas 13½ years is the upper limit of normal development (although 14 years is often used as a simplified limit). **Caucasian** girls have a mean age of menarche later (12.9 years) than **African-American** girls (12.3 years), but this 6-month difference is less than the 1-year difference in the age at onset of other features of puberty between the two groups. A compensation occurs in pubertal development so that those girls who enter puberty at the earliest ages of the normal range take more time to reach menarche, and those who enter at older ages of the normal range progress faster to menarche. This discrepancy is more often found in obese or overweight girls and is attributed to local aromatase activity in adipose tissue.

There is frequent discussion about an earlier age of puberty in children in the United States now than in the past several decades. Data is available from longitudinal studies in the U.S. and abroad. The Bogalusa Heart Study shows that the gap between the age of menarche in **African Americans** and **Caucasians** has widened in the past decades, and this may be due to differing patterns of weight (and fat) gain. However longitudinal study from the Fells Institute shows that those girls who develop early have a tendency to increase their body mass index (BMI) values afterwards more than those who start puberty later. The presence or absence of menarche however, does not affect the interpretation of a BMI value.

Late onset of pubertal development above the upper age limit of normal may indicate hypothalamic, pituitary, or gonadal failure or, alternatively, normal variation (constitutional delay). The time from onset of puberty until adult development is complete is also of importance; significant delays in reaching subsequent stages may indicate any type of hypogonadism.

D. Growth spurt   The striking increase in growth velocity in puberty (pubertal growth spurt) is under complex endocrine control involving thyroid hormone, growth hormone (GH), and estrogen (see Chapter 6, Growth). Hypothyroidism decreases or eliminates the pubertal growth spurt. The amplitude of GH secretion increases in puberty, as does production of insulin-like growth factor (IGF-I); peak serum IGF-I concentrations are reached about 1 year after peak growth velocity, and serum IGF-I levels remain above normal adult levels for up to 4 years thereafter. GH and estrogen are important in the pubertal growth spurt; when either or both are deficient, the growth spurt is decreased or absent. Estrogen indirectly stimulates IGF-I production by increasing the secretion of GH. It also directly stimulates IGF-I production in cartilage. Estrogen is the most important factor in stimulating maturation of the chondrocytes and osteoblasts, ultimately leading to epiphysial fusion. A patient reported with estrogen receptor deficiency was tall, with continued growth past the age of 20 years with a remarkable retardation of skeletal maturation (and decreased bone density). Patients with aromatase deficiency, and therefore, impaired conversion of testosterone to estrogen, also demonstrate diminished advancement of bone age and decreased bone density, as well as continued growth extending into the third decade. With exogenous estrogen administration in the aromatase deficient patients, the bone age advanced, and the bone density increased. Such patients demonstrate the key role played by estrogen in advancing bone age and bringing about the cessation of growth by epiphysial fusion as well as the importance of estrogen in increasing bone density.

In girls, the pubertal growth spurt begins in early puberty and is mostly completed by menarche. In boys, the pubertal growth spurt occurs toward the end of puberty, at an average age 2 years older than in girls. Total height attained during the growth spurt in girls is about 25 cm; in boys, it is about 28 cm. The mean adult height differential of 12 cm between men and women is due in part to heights already attained before onset of the pubertal growth spurt and in part to the height gained during the spurt.

Twin studies from Finland demonstrate that the heritability estimate (proportion of variance of age of onset of puberty attributable to genetic variance) for age at onset of pubertal growth spurt was 0.91, 0.93 for age at peak height velocity, and 0.97 for adult height. In addition, the age at onset of the pubertal growth spurt was negatively associated with BMI in childhood.

Precocious puberty may coexist with GH deficiency: this situation may occur, for example, in a child with a brain tumor causing central precocious puberty who is treated with central nervous system radiation that subsequently causes GH deficiency. The precocious puberty may raise the growth rate to the normal range, thereby cloaking the GH deficiency, but the child does not grow at the excessive rate characteristic of precocious puberty with GH sufficiency.

E. Changes in body composition   Changes in body composition are also prominent during pubertal development.

Prepubertal boys and girls start with equal lean body mass, skeletal mass, and body fat, but at maturity men have approximately 1½ times the lean body mass, skeletal mass, and muscle mass of women, whereas women have twice as much body fat as men. Attainment of peak values of percentage of body fat, lean body mass, and bone mineral density is earlier by several years in girls than in boys, as is the earlier peak of height velocity and velocity of weight gain in girls.

The most important phases of bone accretion occur during infancy and during puberty. Girls reach peak mineralization between 14 and 16 years of age, whereas boys reach a later peak at 17.5 years; both milestones occur after peak height velocity (Figure 15–2). The density of bone is determined by genetic factors; decreased bone mass is found in familial patterns even if subjects are studied before puberty. Patients with delay in puberty for any reason have a significant decrease in bone accretion and a delay in reaching peak bone mineral density, although bone density may later reach normal values in those with constitutional delay in puberty. Moderate exercise increases bone mass, but excessive exercise itself delays puberty; the ultimate outcome of excessive exercise in girls is the combination of exercise-induced amenorrhea, premature osteoporosis, and disordered eating, which is known as the female athletic triad.
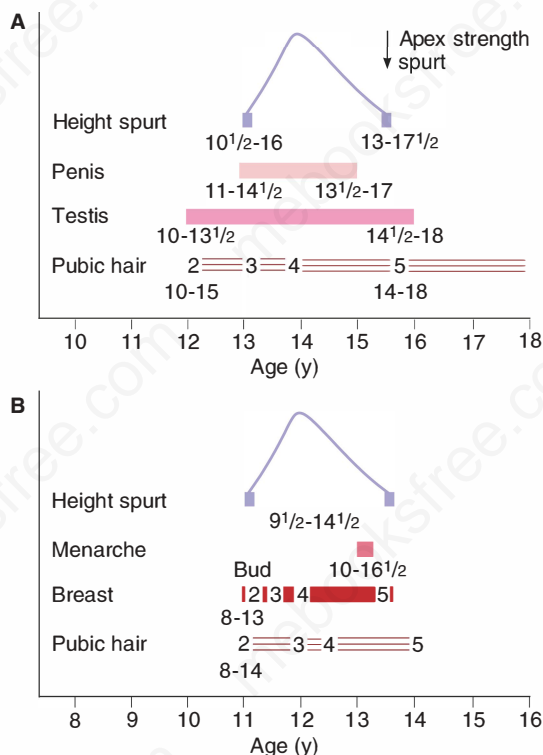
FIGURE 15–2 Sequence of secondary sexual development in British males **(A)** and females **(B)**. The range of ages in Britain is indicated. American boys and girls start pubertal stages earlier than these British children (see text) but the sequence is the same. (Reproduced with permission from Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*. 1970 Feb;45(239):13-23.)

Unfortunately, in the United States, only a minority of adolescents receive the recommended daily allowance of calcium (>1000 mg/d depending on age) and vitamin D, raising concern over future bone health. While a healthy diet is important for all, it is especially important to ensure adequate calcium intake and vitamin D in patients with delayed or absent puberty and in patients treated with gonadotropin-releasing hormone (GnRH) agonists.

F. Other changes of puberty   Other changes that are characteristic of puberty are mediated either directly or indirectly by the change in sex steroids. For example, seborrheic dermatitis may appear, the mouth flora changes, and periodontal disease, rare in childhood, may occur in puberty. Insulin resistance intensifies in normal adolescents as well as those with type 1 diabetes mellitus during puberty. Puberty is a time of extensive change in the anatomy of the brain as pruning of dendritic connections occurs towards the beginning of puberty and myelinization of axons proceeds later; indeed these changes continue until about 25 years of age and are thought to account for the risk-taking behaviors and poor judgment characteristics of early to mid puberty compared to later pubertal individuals who have further maturation of cognitive and executive functions such as judgement. Puberty is also time of the onset of many significant psychiatric disorders including depression. As a reflection of the importance of the changes in brain architecture is the recent discovery of a gene mutation (C4) that inhibits the correct pruning connections in the gray matter and which is associated with the onset of schizophrenia. Sleep patterns change during puberty as, with no light or social cues, adolescents wake up later than in childhood; this change of sleep patterns from a "lark-like" to an "owl-like" pattern of sleep continues until about 19 to 21 years of age.

Endocrine Changes from Fetal Life to Puberty

Pituitary gonadotropin secretion is controlled by the hypothalamus, which releases pulses of GnRH into the pituitary-portal system to reach the anterior pituitary gland by 20 weeks of gestation. Control of GnRH secretion is exerted by a hypothalamic pulse generator in the arcuate nucleus. It is sensitive to feedback control from sex steroids and inhibin, a gonadal protein product that controls the frequency and amplitude of gonadotropin secretion during development in both sexes and during the progression of the menstrual cycle in females (see Chapter 13).

In males, LH stimulates the Leydig cells to secrete testosterone, and FSH stimulates the Sertoli cells to produce inhibin. Inhibin feeds back on the hypothalamic-pituitary axis to inhibit FSH. In females, FSH stimulates the granulosa cells to produce estrogen and the follicles to secrete inhibin, and LH appears to play a minor role in the endocrine milieu until menarche. With menarche, LH triggers ovulation and later stimulates the theca cells to secrete androgens (see Chapter 13). The continuum of development between the fetus and the adult is well illustrated by the changes that occur in the hypothalamic-pituitary-gonadal axis. Gonadotropins are demonstrable in fetal pituitary glands and serum during the first trimester. The pituitary content

of gonadotropins rises to a plateau at midgestation. Fetal serum concentrations of LH and FSH rise to a peak at midgestation and then gradually decrease until term. During the first half of gestation, hypothalamic GnRH content also increases, and the hypophyseal-portal circulation achieves anatomic maturity by midgestation. These data are compatible with a theory of early unrestrained GnRH secretion stimulating pituitary gonadotropin secretion, followed by the appearance of central nervous system factors that inhibit GnRH release and decrease gonadotropin secretion after mid-gestation. Because the male fetus has measurable serum testosterone concentrations but lower serum gonadotropin concentrations than the female fetus, negative feedback inhibition of gonadotropin secretion by testosterone appears to operate after mid-gestation.

A. Changes at birth

At term, serum gonadotropin concentrations are suppressed in the neonate, but with postnatal clearance of high circulating estrogen concentrations, negative inhibition is reduced and postnatal peaks of serum LH and FSH are measurable for several months to up to a few years after birth. Serum testosterone concentrations may be increased to midpubertal levels during the several months after birth in normal males and serum estradiol rises in pulses for the first 2 years in infant girls; this period is considered the mini-puberty of infancy. These peaks of gonadotropins and sex steroids in normal infants complicate the diagnosis of central precocious puberty at these youngest ages, because it is difficult to decide whether to attribute the gonadotropin and sex steroid peaks to central precocious puberty or to normal physiology (the mini-puberty of infancy). While episodic peaks of serum gonadotropins may occur until 2 years of age, serum gonadotropin concentrations are low during later years in normal childhood.

B. The juvenile pause or the mid-childhood nadir of gonadotropin secretion

While serum gonadotropin concentrations are low in midchildhood, sensitive assays indicate that pulsatile secretion occurs and that the onset of puberty is heralded by an increase in amplitude of secretory events and a change in frequency. Twenty-four hour mean concentrations of LH, FSH, and testosterone rise measurably within 1 year before the development of physical pubertal changes. Patients with gonadal failure—such as those with the syndrome of gonadal dysgenesis (Turner syndrome)—demonstrate an exaggeration of the normal pattern of gonadotropin secretion, with exceedingly high concentrations of serum LH and FSH during the first several years of life. Such patients demonstrate that negative feedback inhibition is active during childhood; without sex steroid or inhibin secretion to exert inhibition, serum gonadotropin values are greatly elevated. During midchildhood, normal individuals and patients with primary hypogonadism have lower serum gonadotropin levels than they do in the neonatal period, but the range of serum gonadotropin concentrations in hypogonadal patients during midchildhood is still higher than that found in healthy children of the same age. The decrease in serum gonadotropin concentrations in primary agonadal children during midchildhood is incompletely understood, but has been attributed to an increase in the central nervous

system inhibition of gonadotropin secretion during these years. This inhibition is mediated by gamma aminobutyric acid (GABA) and other inhibitory neuropeptides. The recent discovery of the *MRKN3* (makorin ring finger protein 3) gene which is considered a "brake" on puberty adds considerably more understanding of the juvenile pause; when this gene is silenced, precocious puberty occurs. Thus, the juvenile pause in normal children and those with primary gonadal failure appears to be due to central nervous system restraint of GnRH secretion.

C. Peripubertal gonadotropin increase

Prepubertal children demonstrate a circadian rhythm of LH and FSH secretion of low amplitude with a pattern of low levels of sex steroid secretion lagging behind the gonadotropin rhythm. The delay is presumably due to the time necessary for biosynthesis and secretion of sex steroids. Thus, the changes that occur at puberty do not arise *de novo* but are based on preexisting patterns of endocrine secretion. In the peripubertal period, endogenous GnRH secretion increases in amplitude and frequency during the early hours of sleep and serum testosterone and estrogen concentrations rise several hours later, again suggesting that biosynthesis or aromatization occurs during the period of delay—a pattern that differs from the prepubertal period mainly in the increased amplitude of the secretion encountered in puberty. As puberty progresses in both sexes, the peaks of serum LH and FSH occur more often during waking hours; and, finally, in late puberty, the peaks occur at all times, eliminating the diurnal variation.

During the peripubertal period of endocrine change prior to secondary sexual development, gonadotropin secretion becomes less sensitive to negative feedback inhibition. Before this time, a small dose of exogenous sex steroids virtually eliminates gonadotropin secretion, whereas afterward a far larger dose is required to suppress serum FSH and LH. Thus, an equilibrium balancing pubertal concentrations of gonadotropins and sex steroids is reached.

The switch that triggers the onset of puberty is unknown, but several neurotransmitters are invoked in the process, including GABA and *N*-methyl-D-aspartate. *KISS1*, a human metastasis suppressor gene at locus 19p13.3, which codes for metastin (also known as kisspeptin), a 145 amino acid peptide, is an important agent in the process. Metastin is the endogenous agonist for GPR54, a Gq/11-coupled receptor of the rhodopsin family (metastin receptor), which is found in the brain, mainly in the hypothalamus and basal ganglia, as well as in the placenta. *KISS1* mRNA levels increase with puberty in intact male and female monkeys from the juvenile to midpubertal stages. Furthermore, administration of KISS1 via intracerebral catheters to GnRH-primed juvenile female rhesus monkeys stimulates GnRH release, release that is abolished by infusion of GnRH antagonist. Thus, it is postulated that KISS1 signaling through the GPR54 receptor of primate hypothalamus may be activated at the end of the juvenile pause and contribute to the pubertal resurgence of pulsatile GnRH release of puberty, serving as a trigger for the onset of puberty. A decrease in expression of *MKRN3*, a gene which is considered a brake on pubertal development, occurs in the pubertal period; decreased expression of this imprinted gene is passed through the father and leads to precocious puberty.

Highly sensitive *sandwich* assays (immunoradiometric [IRMA]) and immunochemiluminometric assays (ICMA), "third-generation assays" are available for gonadotropin determination. They can be used to indicate the onset of pubertal development solely with basal samples usually without the necessity for GnRH testing (see GnRH and GnRH agonist testing in the Central Precocious Puberty section). Basal LH values in the pubertal range (different laboratory reference ranges differ) determined by third-generation assays in random blood samples are highly predictive of elevated peak GnRH-stimulated LH and therefore indicate the onset of central precocious puberty or normal puberty. These third-generation assays further reflect the remarkable logarithmic increase in spontaneous LH secretion in the latest stages of prepuberty and earliest stages of puberty as the testicular volume increases from 1 to 2 mL to 10 mL; these increases in serum LH are far greater proportionately than those found in the last stages of pubertal development. The magnitude of increase in serum testosterone is also greater in the early stages of male puberty and correlates with the increase in serum LH during this same period of early pubertal development. Note that standard LH and FSH assays are not sensitive enough to detect the changes seen with the ultrasensitive assays. Furthermore, the laboratory must have pubertal standard values available to permit accurate interpretation of the measurements. Since both types of assays are available commercially, care in ordering is needed to ensure a pediatric assay is chosen.

D. Sex steroid secretion   Sex steroid secretion is correlated with the development of gonadotropin secretion. During the postnatal period of increased episodic gonadotropin secretion, plasma concentrations of gonadal steroids are episodically elevated. This indicates the potential for secretory activity in the neonatal gonad. Later, when gonadotropin secretion decreases in midchildhood, gonadal activity decreases, but the testes can still be stimulated by LH or hCG and the ovaries by FSH, with resulting secretion of gonadal steroids. An ultrasensitive estradiol assay demonstrates higher values of serum estradiol in prepubertal girls than prepubertal boys, indicating basal ovarian activity during the juvenile pause. With the onset of puberty, serum gonadal steroid concentrations progressively increase. While sex steroids are secreted in a diurnal rhythm in early puberty, as are gonadotropins, sex steroids are bound to sex hormone–binding globulin (SHBG), so that the half-life of sex steroids is longer than that of gonadotropins. Thus, random daytime measurements of serum sex steroids are more helpful in determining pubertal status than random measurements of serum gonadotropins but still are not infallible, for example, early pubertal girls may or may not have pubertal values of estradiol on any given sample.

Most (97%-99%) of the circulating estradiol and testosterone is associated with SHBG. The free hormone is the active fraction, but SHBG modulates the activity of the total testosterone and estradiol. Prepubertal boys and girls have equal concentrations of SHBG, but because testosterone decreases SHBG and estrogen increases SHBG, adult males have only half the concentration of SHBG compared with adult females. Thus, lower SHBG levels amplify androgen effects in men. While adult men have 20 times the amount of plasma testosterone that adult women have, adult men have 40 times the amount of free testosterone that adult women have.

E. GnRH stimulation   The use of exogenous GnRH has further clarified the pattern of pubertal development. When GnRH or GnRH agonist is administered to children less than 2 years of age, pituitary secretion of LH and FSH increases markedly. During the juvenile pause, the period of low basal gonadotropin secretion (after age 2 until the peripubertal period), exogenous GnRH has less effect on LH release. By the peripubertal period, GnRH or GnRH agonist induces a greater rise in LH concentrations in boys and girls, and this response continues until adulthood. There is no significant change in the magnitude of FSH secretion after GnRH with the onset of puberty, though females at all ages release more FSH than males.

The rise of LH after GnRH is used as a diagnostic test to differentiate central precocious puberty from the prepubertal state. Values are laboratory specific, but it has been reported that a basal LH value over 0.3 or a peak LH over 5.5 U/L one hour after 10 µg/kg subcutaneous GnRH administration indicates the onset of puberty with 96% specificity in girls. More specificity is reported in boys with a 4-hour peak LH value over 14.8 to 15 units per liter after 10 µg/kg of leuprolide acetate.

Gonadotropins are released in secretory spurts in response to endogenous GnRH, which itself is secreted episodically about every 90 to 120 minutes in response to a central nervous system *pulse generator*. GnRH can be administered to patients in episodic boluses by a programmable pump that mimics the natural secretory episodes. A prepubertal subject without significant gonadotropin peaks demonstrates the normal pubertal pattern of episodic secretion of gonadotropins after only a few days of exogenously administered GnRH boluses. Hypogonadotropic patients, who in the basal state do not have normal secretory episodes of gonadotropin release, may be stimulated to a pattern of normal adult episodic gonadotropin secretion by this method of pulsatile GnRH administration.

This phenomenon is used in clinical practice to bring about ovulation or spermatogenesis to foster fertility. Varying the timing of pulsatile exogenous GnRH administration can regulate the ratio of FSH to LH just as the FSH to LH ratio changes as the frequency of endogenous hypothalamic GnRH release shifts during the menstrual cycle and to some degree in puberty naturally. Increasing the frequency of GnRH pulses increases the LH-FSH ratio; an increased ratio is characteristic of midcycle and peripubertal dynamics. Alternatively, if GnRH is administered continuously rather than in pulses or if long-acting superactive analogs of GnRH are given, a brief period of increased gonadotropin secretion is followed by LH and FSH suppression (see later). This phenomenon is responsible for the therapeutic effects of GnRH analogs in conditions such as central precocious puberty.

F. Leptin and puberty   Leptin is a hormone produced in adipose cells that suppresses appetite through interaction with its receptor in the hypothalamus. Leptin plays a major role in pubertal development in mice and rats. Genetically leptin-deficient mice (ob/ob) do not initiate puberty. Leptin replacement promotes pubertal development in this mouse, and leptin

administration causes an immature normal mouse to progress through puberty. A leptin-deficient girl at 9 years of age was remarkably obese and had a bone age of 13 years (usually appropriate for the onset of normal puberty) but no significant gonadotropin pulses and no physical evidence of pubertal development. With leptin treatment, gonadotropin peaks appeared, and pubertal development followed. Individuals with leptin receptor deficiency also have disordered puberty. Longitudinal studies demonstrate that leptin increases in girls during puberty in synchrony with the increase in fat mass, while leptin decreases in boys, with increased lean body mass and decreased fat mass, in relation to testosterone production. While leptin does not appear to trigger the onset of puberty in normal adolescents, changes in plasma leptin concentrations may accompany pubertal changes rather than cause them. Leptin is a necessary component of pubertal development in human beings but not a major stimulant to this development; leptin is necessary but not sufficient for pubertal development.

Ovulation and Menarche

The last stage in hypothalamic-pituitary development is the onset of positive feedback, leading to ovulation and menarche. The ovary contains a paracrine system that regulates follicular atresia or development; it is only in the last stages of puberty that gonadotropins come into play in the maturation of the follicle. After midpuberty, estrogen in the appropriate amount at the appropriate time can stimulate gonadotropin release, whereas higher doses of estrogen can suppress gonadotropin secretion. The frequency of pulsatile GnRH release increases during the late follicular phase of the normal menstrual cycle and raises the ratio of LH to FSH secretion. This stimulates the ovary to produce more estrogen and leads to the midcycle LH surge that causes ovulation. Administration of pulsatile GnRH by programmable pump can be used to bring about fertility in patients with hypothalamic GnRH deficiency by mimicking this natural pattern as noted earlier.

However, even if the midcycle surge of gonadotropins is present, ovulation may not occur during the first menstrual cycles; 90% of menstrual cycles are anovulatory in the first year after menarche, and it is not until 4 to 5 years after menarche that the percentage of anovulatory cycles decreases to less than 20%. This high prevalence of anovulatory periods may be due to unrecognized PCOS in the population studied sound rather than a normal developmental phenomenon. However, some of the first cycles after menarche may be ovulatory, and fertility is possible in the first cycle.

Adrenarche

Although the hypothalamic-pituitary axis has been well characterized in recent years, our understanding of the mechanism of control of adrenal androgen secretion is still somewhat rudimentary. The adrenal cortex normally secretes the weak androgens dehydroepiandrosterone (DHEA), its sulfate, dehydroepiandrosterone sulfate (DHEAS), and androstenedione in increasing amounts beginning at about 6 to 7 years of age in girls and 7 to 8 years of age in boys (Table 15–1). A continued rise in adrenal androgen secretion persists until late puberty. Thus, adrenarche (the secretion of adrenal androgens) occurs years before gonadarche (the secretion of gonadal sex steroids). Patients with Addison disease, who do not secrete adrenal androgens, and patients with premature adrenarche, who secrete increased amounts of adrenal androgens at an early age, usually enter gonadarche at a normal age. This suggests that age at adrenarche does not significantly influence age at gonadarche. Furthermore, patients treated with a GnRH agonist to suppress gonadotropin secretion progress through adrenarche despite their suppressed gonadarche.

Miscellaneous Metabolic Changes

The onset of puberty is associated with many changes in laboratory values that are either directly or indirectly caused by the rise of sex steroid concentrations and careful attention to the reference range of laboratory data during puberty is important. Thus, in boys, hematocrit rises, and high-density lipoprotein (HDL) concentrations fall as a consequence of increasing testosterone. In both boys and girls, alkaline phosphatase rises normally during the pubertal growth spurt (often this rise is incorrectly interpreted as evidence for a tumor or liver abnormality). Serum IGF-I concentrations rise with the pubertal growth spurt, but IGF-I is more closely correlated with sex steroid concentration than with growth rate. IGF-I levels reach a peak 1 year after peak growth velocity is reached and remain elevated for 4 years thereafter, even though growth rate is decreasing. Prostate-specific antigen (PSA) is measurable after the onset of puberty in boys and provides another biochemical indication of pubertal onset.

DELAYED PUBERTY OR ABSENT PUBERTY (SEXUAL INFANTILISM)

Any girl of 13 or boy of 14 years (13.5 years in some sources) of age without signs of pubertal development falls more than 2.5 SD above the mean and is considered to have delayed puberty (Table 15–2). By this definition, 0.6% of the healthy population is classified as having delay in growth and adolescence. These normal patients need reassurance rather than treatment and ultimately progress through the normal stages of puberty, albeit later than their peers. The examining physician must make the sometimes difficult decision about which patients older than these guidelines are constitutionally delayed and which have organic disease.

Constitutional Delay in Growth and Adolescence

A patient with delayed onset of secondary sexual development, whose stature is shorter than that of age-matched peers but who consistently maintains a normal growth velocity for bone age and whose skeletal development is delayed more than two SD from the mean, is likely to have constitutional delay in puberty (Figure 15–3). These patients are at the older end of the normal distribution curve describing the age at onset of puberty. A family history

TABLE 15–1 Pubertal hormone reference ranges.

LH

Puberty

Tanner Stage	Age (years)	Range (mIU/mL)		Tanner Stage	Age (years)	Range (mIU/mL)
Male				**Female**		
1	<9.8	0.02-0.3		1	<9.2	0.02-0.18
2	9.8-14.5	0.2-4.9		2	9.2-13.7	0.02-4.7
3	10.7-15.4	0.2-5.0		3	10.0-14.4	0.10-12.0
4-5	11.8-17.3	0.4-7.0		4-5	10.7-18.6	0.4-11.7

FSH

Puberty

Tanner Stage	Age (years)	Range (pg/dL)	Mean (µg/dL)	Tanner Stage	Age (years)	Range (pg/dL)	Mean (µg/dL)
Male				**Female**			
1	<9.8	0.26-3.0	0.98	1	<9.2	1.0-4.2	2.1
2	9.8-14.5	1.8-3.2	2.5	2	9.2-13.7	1.0-10.8	4.0
3	10.7-15.4	1.2-5.8	2.9	3	10.0-14.4	1.5-12.8	5.1
4	11.8-16.2	2.0-9.2	4.4	4	10.7-15.6	1.5-11.7	6.4
5	12.8-17.3	2.6-11.0	6.1	5	11.8-18.6	1.0-9.2	4.9

DHEAS

Puberty

Tanner Stage	Age (years)	Range (µg/dL)		Tanner Stage	Age (years)	Range (µg/dL)
Male				**Female**		
1	<9.8	13-83		1	<9.2	19-44
2	9.8-14.5	42-109		2	9.2-13.7	34-129
3	10.7-15.4	48-200		3	10.0-14.4	32-226
4	11.8-16.2	102-385		4	10.7-15.6	58-260
5	12.8-17.3	120-370		5	11.8-18.6	44-248

Estradiol

Puberty

Tanner Stage	Age (years)	Range (µg/dL)		Tanner Stage	Age (years)	Range (µg/dL)
Male				**Female**		
1	<9.8	5.0-11		1	<9.2	5.0-20
2	9.8-14.5	5.0-16		2	9.2-13.7	10-24
3	10.7-15.4	5.0-25		3	10.0-14.4	7.0-60
4	11.8-16.2	10-36		4	10.7-15.6	21-85
5	12.8-17.3	10-36		5	11.8-18.6	34-170

Testosterone

Puberty

Tanner Stage	Age (years)	Range (pg/dL)	Mean (µg/dL)	Tanner Stage	Age (years)	Range (pg/dL)	Mean (µg/dL)
Male				**Female**			
1	<9.8	<2.5-10	4.9	1	<9.2	<2.5-10	4.9
2	9.8-14.5	18-150	42	2	9.2-13.7	7-28	18
3	10.7-15.4	100-320	190	3	10.0-14.4	15-35	25
4	11.8-16.2	200-620	372	4	10.7-15.6	13-32	22
5	12.8-17.3	350-970	546	5	11.8-18.6	20-38	28

Estradiol, ultrasensitive by HPLC MS/MS, in cycling females:

Follicular stage 39-375 pg/mL

Midcycle stage 94-762 pg/mL

Luteal stage 48-440 pg/mL

LH-pediatric is an immunometric (sandwich) electrochemiluminescent (ECL) ultrasensitive assay. FSH-pediatric is an immunometric (sandwich) chemiluminescent ultrasensitive ELISA (C-ELISA) assay.

Testosterone by HPLC MS/MS.

Estradiol, ultrasensitive is a tandem mass spectrometry by HPLC MS/MS assay.

DHEA-S is a competitive chemiluminescent immunoassay.

Data from Styne DM. *Pediatric Endocrinology, A Clinical Handbook*. Switzerland: Springer International Publishing; 2016.

TABLE 15–2 Classification of delayed puberty.

Constitutional delay in growth and adolescence
Hypogonadotropic hypogonadism
Central nervous system disorders
Congenital disorders of the hypothalamus or pituitary
Tumors
Other acquired disorders
Infection
Trauma
Irradiation
Defects of the hypothalamic-pituitary axis
Isolated gonadotropin deficiency
Kallmann syndrome
Gonadotropin deficiency with normal sense of smell
Multiple pituitary hormonal deficiencies
Miscellaneous disorders
Syndromes
Chronic disease
Weight loss
Anorexia nervosa
Increased physical activity in female athletes
Hypothyroidism
Genetic syndromes
Prader-Willi syndrome
Bardet-Biedl syndrome
Hypergonadotropic hypogonadism
Male phenotype
Klinefelter syndrome
Other forms of primary testicular failure (including chemotherapy)
Enzymatic defects of androgen production
Anorchia or cryptorchism
Female phenotype
Turner syndrome
Other forms of primary ovarian failure (including chemotherapy)
Pseudo–Turner syndrome
Noonan syndrome
XX and XY gonadal dysgenesis

of a similar pattern of development in a parent or sibling supports the diagnosis. A mother or sister who starts menarche after 15 years of age or a father or brother who started shaving or experienced considerable growth after high school are supportive historical details. The subject is usually thin as well. In many cases, even if they show no physical signs of puberty at the time of examination, the initial elevation of gonadal sex steroids has already begun, and their basal LH concentrations measured by ultrasensitive third-generation assays or their plasma LH response to exogenous GnRH or GnRH agonist is pubertal. In boys, an 8 AM serum testosterone value above 20 ng/dL (0.7 mmol/L) also indicates that secondary sexual development will commence within a period of 15 months.

In some cases, observation for endocrine or physical signs of puberty must continue for a period of months or years before the diagnosis is made. Generally, signs of puberty appear after the patient reaches a skeletal age of 11 years (girls) or 12 years (boys), but there is great variation. Patients with constitutional delay in adolescence almost always manifest secondary sexual development by 18 years of chronologic age, although there is one reported case of spontaneous puberty occurring at 25 years of age. Reports of patients with Kallmann syndrome and others with constitutional

delay in puberty within one family suggest a possible relationship between the two conditions (see later and Chapter 4). Adrenarche is reported to be delayed—along with gonadarche—in constitutional delay in puberty.

Hypogonadotropic Hypogonadism

The absent or decreased ability of the hypothalamus to secrete GnRH or of the pituitary gland to secrete LH and FSH characterizes hypogonadotropic hypogonadism. This classification suggests a long-term condition requiring replacement therapy. If the pituitary deficiency is limited to gonadotropins, a patient's stature is usually close to average height for age until the usual age of the pubertal growth spurt, in contrast to shorter patients with constitutional delay. Bone age is usually not delayed in childhood but does not progress normally after the patient reaches the age at which sex steroid secretion usually stimulates further maturation of the skeleton. Patients may reach taller stature than expected for family. However, if GH deficiency accompanies gonadotropin deficiency, severe short stature will result.

A. Central nervous system disorders

1. **Tumors**—A tumor involving the hypothalamus or pituitary gland can interfere with hypothalamic-pituitary-gonadal function as well as control of GH, adrenocorticotropic hormone (ACTH), thyrotropin (TSH), prolactin (PRL), and vasopressin secretion. Thus, delayed puberty may be a manifestation of a central nervous system tumor accompanied by any or all of the following: GH deficiency, secondary hypothyroidism, secondary adrenal insufficiency, hyperprolactinemia, and diabetes insipidus. The combination of anterior and posterior pituitary deficiencies acquired after birth makes it imperative that a hypothalamic-pituitary tumor be eliminated as the cause before assigning another diagnosis.

   Craniopharyngioma is rare but is the most common type of hypothalamic-pituitary tumor leading to delay or absence of pubertal development. This neoplasm originates in Rathke's pouch but may develop into a suprasellar tumor. The peak age incidence of craniopharyngioma is between 6 and 14 years. Presenting symptoms may include headache, visual deficiency, growth failure, polyuria, and polydipsia; presenting signs may include visual defects such as bitemporal hemianopsia due to optic atrophy, or, alternatively, papilledema. Clinical manifestations and laboratory evaluation may reflect endocrinopathies (found in 80% including GH axis abnormalities in 75%, hyperprolactinemia in 20%, hypothyroidism in 25%, adrenal insufficiency in 25%, diabetes insipidus (DI) in 9% to 17%, as well as gonadotropin deficiency in 40% in one large series). Bone age is often delayed at the time of presentation.

   Calcification in the suprasellar region is the hallmark of craniopharyngiomas; 80% of cases have calcifications on lateral skull x-ray, and a higher percentage show this on computed tomography (CT); while calcifications cannot be seen on standard magnetic resonance imaging (MRI), susceptibility weighted imaging (SWI) allows an MRI to reveal calcification. The tumor often presents a cystic appearance on CT or MRI scan and at the time of surgery may contain dark, cholesterol-laden fluid. The rate of growth of craniopharyngiomas is quite variable—some are indolent and some are quite aggressive. Small intrasellar tumors may be resected by transsphenoidal
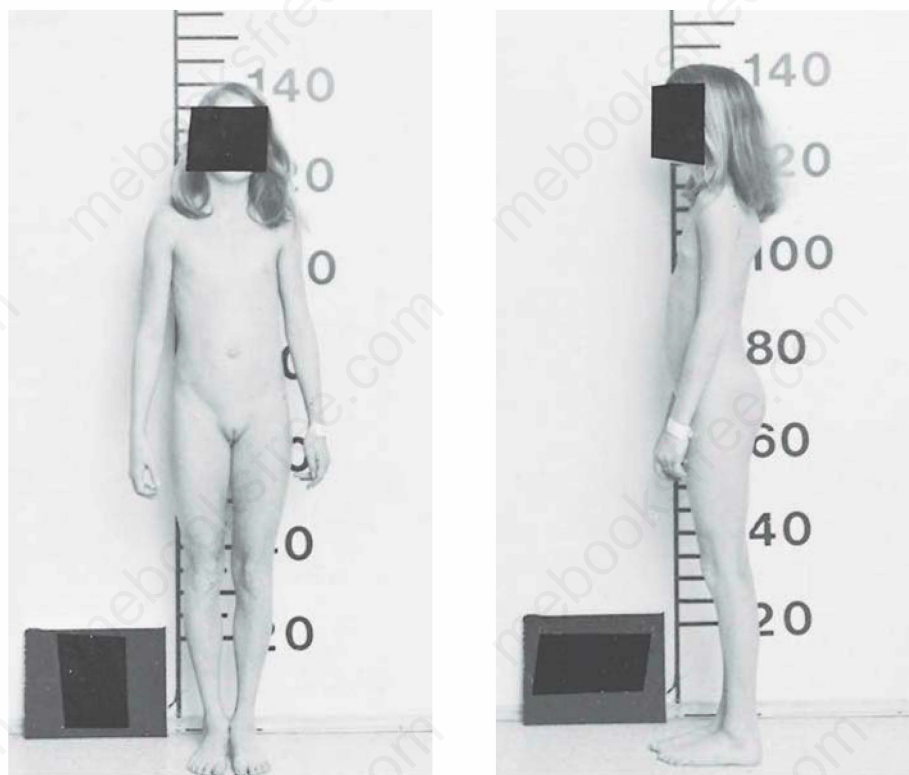
FIGURE 15–3 Girl, 13-4/12 years old, with constitutional delay in growth and puberty. History revealed a normal growth rate but short stature at all ages. Physical examination revealed a height of 138 cm (−4.5 SD) and a weight of 28.6 kg (−3 SD). The patient recently had early stage 2 breast development, with 1 cm of glandular tissue on the right breast and 2 cm on the left breast. The vaginal mucosa was dulled, and there was no pubic hair. Karyotype was 46,XX. Bone age was 10 years. After administration of GnRH, LH and FSH rose in a pubertal pattern. Estradiol was 40 pg/mL. She has since spontaneously progressed through further pubertal development. (Reproduced with permission from Styne DM, Kaplan SL. Normal and abnormal puberty in the female. *Pediatr Clin North Am*. 1979 Feb;26(1):123-148.)

surgery; larger ones require partial resection and radiation therapy (see Chapter 4). Recurrence of this tumor after apparent complete removal is reported and lends support to the use of radiation therapy in addition to surgery.

Extrasellar tumors that involve the hypothalamus and cause sexual infantilism include germinomas, gliomas (possibly associated with neurofibromatosis), and astrocytomas (see Chapter 4). However, depending on endocrine secretory activity or location, these tumors may alternatively cause central precocious puberty. Intrasellar tumors such as chromophobe adenomas are quite rare in children compared with adults. Hyperprolactinemia—with or without a diagnosed microadenoma or galactorrhea—may delay the onset or progression of puberty; with dopaminergic therapy to decrease PRL concentrations, puberty progresses.

2. **Other acquired central nervous system disorders**—These disorders may lead to hypothalamic-pituitary dysfunction. Granulomatous diseases such as Hand-Schüller-Christian disease or histiocytosis X, when involving the hypothalamus, most frequently lead to diabetes insipidus, but any other hypothalamic defect may also occur, including GH or gonadotropin deficiency. Tuberculous or sarcoid granulomas, other postinfectious inflammatory lesions, vascular lesions, and trauma more rarely cause hypogonadotropic hypogonadism.

3. **Developmental defects**—Developmental defects of the central nervous system may cause hypogonadotropic hypogonadism or other types of hypothalamic dysfunction. Cleft palate or other midline anomalies may also be associated with hypothalamic dysfunction. Optic dysplasia is associated with small, hypoplastic optic disks and, in some patients, absence of the septum pellucidum on CT or MRI (septo-optic dysplasia); associated hypothalamic deficiencies are often present (see Chapters 4 and 6). This may be caused by mutations in *Hesx-1 homeodomain,* (*HESX-1* OMIM*601802). *Homeobox* gene expressed in ES cells (HESX-1 OMIM#182230, septooptic dysplasia).

Optic hypoplasia or dysplasia must be differentiated from optic atrophy; optic atrophy implies an acquired condition and may indicate a hypothalamic-pituitary tumor. Both anterior and/or posterior pituitary deficiencies may occur with either congenital midline defects or acquired hypothalamic-pituitary defects. Early-onset GH deficiency or early onset of a combination of anterior and posterior pituitary deficiencies suggests a congenital defect. Patients who have isolated deficiency of gonadotropins tend to be of normal height until the teenage years but lack a pubertal spurt. They have eunuchoid proportions of increased span for height and decreased upper to lower segment ratios. Their skeletal development is delayed for chronologic age during the teenage years; they continue to grow after an age when normal adolescents stop growing. Adult height is often increased in hypogonadotropic hypogonadal individuals.

4. Radiation therapy—Central nervous system radiation therapy involving the hypothalamic-pituitary area can lead to hypogonadotropic hypogonadism with onset at 6 to 18 months (or longer) after treatment. GH is more frequently affected than gonadotropin secretion, and GH deficiency occurs with exposure to an 18-Gy dose. Other hypothalamic deficiencies such as gonadotropin deficiency, hypothyroidism, and hyperprolactinemia occur more often with higher doses of radiation.

B. Isolated gonadotropin deficiency **Kallmann syndrome 1** is the most common genetic form of isolated gonadotropin deficiency (Figure 15–4). Gonadotropin deficiency in these patients is associated with hypoplasia or aplasia of the olfactory lobes and olfactory bulb causing hyposmia or anosmia. Remarkably, patients may not notice that they have no sense of smell, although olfactory testing reveals the finding. GnRH-containing neurons fail to migrate from the olfactory placode (where they originate) to the medial basal hypothalamus in Kallmann syndrome. This is a familial syndrome of variable manifestations in which anosmia may occur with or without hypogonadism in a given member of a kindred. **X-linked Kallmann syndrome** is due to gene deletions in the region of Xp22.3 (OMIM#308700. KAL1), causing the absence of the *KAL1* gene which encodes for anosmin, an adhesion molecule that plays a key role in the migration of GnRH neurons and olfactory nerves to the hypothalamus.

There is an association of Kallmann syndrome 1 with X-linked ichthyosis due to steroid sulfatase deficiency, developmental delay, and chondrodysplasia punctata, probably due to a microdeletion. Associated abnormalities in Kallmann syndrome 1 may affect the kidneys and bones, and patients may have undescended testes, gynecomastia, and obesity. Mirror hand movements (bimanual synkinesia) are reported, with MRI evidence of abnormal development of the corticospinal tract. Adult height is normal, although patients are delayed in reaching adult height. **Kallmann syndrome 2** (OMIM#147950) is inherited in an autosomal dominant pattern and is due to a mutation in the *FGFR1* (*fibroblast growth factor receptor 1* OMIM*136350) gene. **Kallmann syndrome 3** (OMIM#244200) exhibits an autosomal recessive pattern and appears to be related to mutations in *PROKR2* and *PROK2*, encoding prokineticin receptor-2 and prokineticin-2, respectively. FGF8 may also be involved.

While kisspeptin (OMIM*603286, KISS1 METASTASIS SUPPRESSOR; KISS1) and GRP54 (G protein–coupled receptor 54 (OMIM*604161, KISS1 RECEPTOR; KISS1R)) play roles of importance in the onset of puberty, only rare patients have been reported with defects in the kisspeptin-GPR54 axis due to a loss of function mutation in the gene for the GRP54 receptor (OMIM#614842). The first human beings with defects in these molecules were reported in several kindreds, some consanguineous,
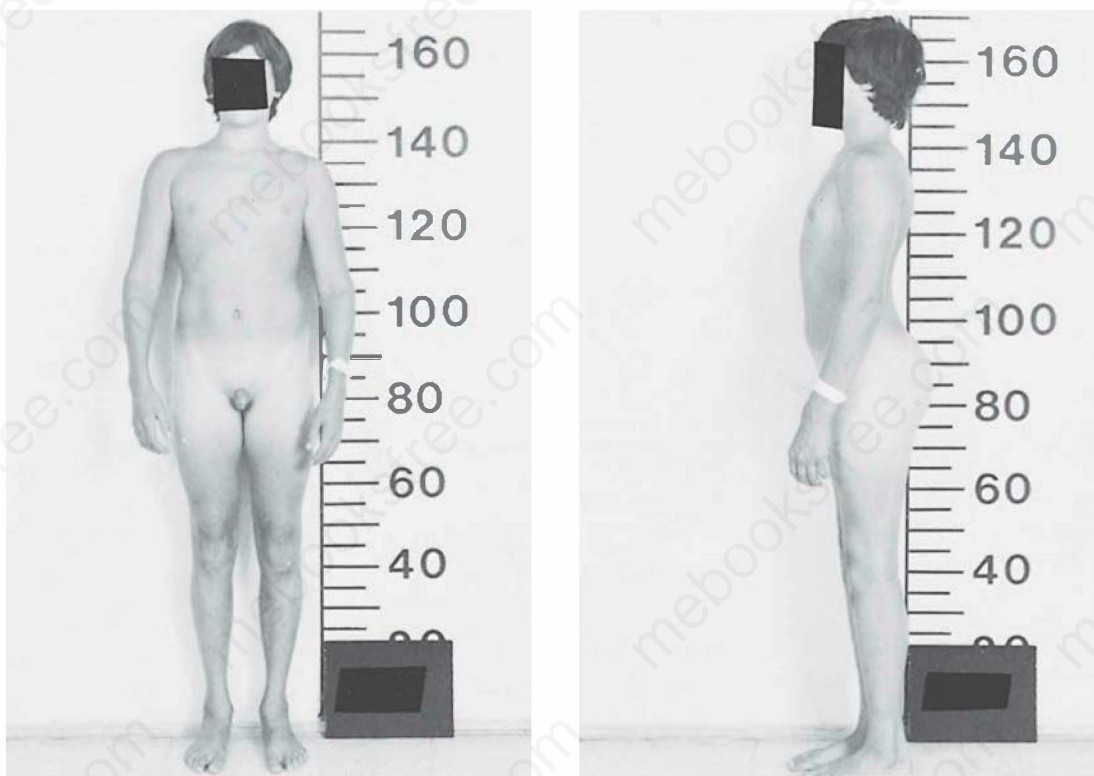
FIGURE 15–4 Boy, 15-10/12 years old, with Kallmann syndrome. His testes were originally undescended, but they descended into the scrotum after human chorionic gonadotropin treatment was given. His height was 163.9 cm (−1.5 SD), and the US-LS ratio was 0.86 (eunuchoid). The penis measurement was prepubertal at 6.3 cm × 1.8 cm. Each testis was 1 cm × 2 cm, prepubertal. Plasma LH was not detectable and rose little after administration of 100 μg of GnRH; FSH rose minimally. Testosterone did not change from 17 ng/dL. He had no ability to smell standard odors. (Reproduced with permission from Yen SSC, Jaffe RB. *Reproductive Endocrinology*. Philadelphia: WB Saunders; 1978.)

presenting with hypogonadotropic hypogonadism. Other cases of hypogonadotropic hypogonadism may occur sporadically or in an autosomal recessive pattern without anosmia; there are 22 types of HH with or without anomia listed on the Online Mendelian Inheritance in Man (OMIM) site. The *GnRH* gene would seem a likely candidate for a mutation as the cause of the condition, but while mutations of the *GnRH receptor* gene (OMIM#146110), *GnRHR*, were noted years ago, not until 2009 were mutations in the *GnRH1* gene demonstrated (OMIM#614841). Other mutations causing hypogonadotropic hypogonadism without anosmia include the *SF-1* (*steroidogenic factor 1*), *HESX-1* (*Hesx-1 homeodomain*), *LHX3* (OMIM*600577) (*LIM homeobox gene 3*), *PROP-1* (*prophet of PIT1*) genes. X-linked congenital adrenal hypoplasia is associated with hypogonadotropic hypogonadism (*DAX1* OMIM#300200). Adrenal hypoplasia, congenital AHC, glycerol kinase deficiency, and muscular dystrophy have also been linked to this syndrome. The gene locus is at Xp21.3-p21.2 and involves a mutation in the *DAX1* gene in many but not all patients. An autosomal recessive form of congenital adrenal hypoplasia is reported. Some hypogonadal patients lack only LH secretion and have spermatogenesis without testosterone production (fertile eunuch syndrome); others lack only FSH (see Chapters 12 and 13).

While a specific genetic diagnosis for hypogonadotropic hypogonadism suggests a permanent defect, long-term studies do not always bear this out. Men with a variety of mutations followed for years are reported to revert to normal or near-normal gonadotropin function in 10% to 20% of cases. Thus, long-term follow-up is indicated.

C. Idiopathic hypopituitary dwarfism (growth hormone deficiency in the absence of defined anatomic or organic defects)

Patients with congenital GH deficiency have early onset of growth failure (Figure 15–5); this feature distinguishes them from patients with GH deficiency due to hypothalamic tumors, who usually have late onset of growth failure. Even without associated gonadotropin deficiency, untreated GH-deficient patients often have delayed onset of puberty associated with their delayed bone ages. With appropriate human growth hormone (hGH) therapy, however, onset of puberty occurs at a normal age. Patients who have combined GH and gonadotropin deficiency do not undergo puberty even when bone age reaches the pubertal stage. Idiopathic hypopituitarism is usually sporadic, but may follow an autosomal recessive or X-linked inheritance pattern due to one of the gene defects listed earlier.

The syndrome of microphallus (penile length <2 cm at birth due to congenital gonadotropin or GH deficiency) and neonatal hypoglycemic seizures (due to congenital ACTH deficiency or GH deficiency) must be diagnosed and treated early to avoid central nervous system damage due to hypoglycemia (these findings may also occur in septo-optic dysplasia). Patients with this syndrome do not undergo spontaneous pubertal development. Testosterone in low doses (testosterone enanthate, 25 mg intramuscularly every month for three doses) can increase the size of the penis in infants diagnosed with congenital hypopituitarism without significantly advancing the bone age. Males with isolated
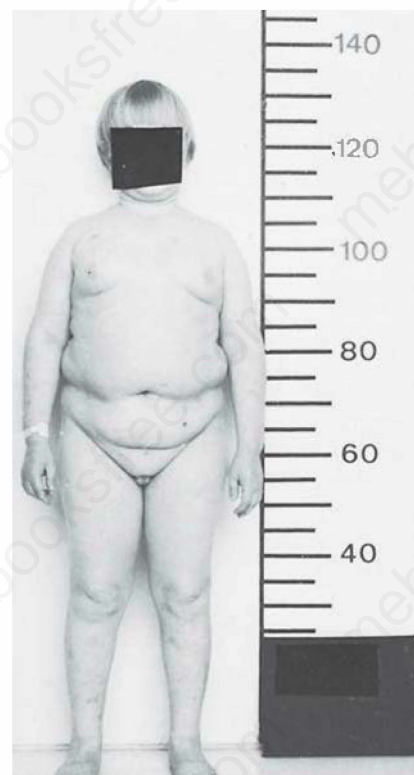
FIGURE 15–5 Twenty-year-old male with congenital deficiency of GRF, GnRH, TRF, and CRF. Height was 8 SD below the mean, and the phallus was 2 cm × 1 cm. Bone age was 10 years, and the sella turcica was small on lateral skull x-ray. LH rose minimally from a low basal value after administration of 100 μg of GnRH. Testosterone was virtually undetectable and did not rise after administration of GnRH. (Reproduced with permission from Yen SSC, Jaffe RB. *Reproductive Endocrinology*. Philadelphia: WB Saunders; 1978.)

GH deficiency can also have microphallus; the penis enlarges with hGH therapy in these patients. It is important to note that microphallus due to hypopituitarism may be successfully treated with testosterone leading to adult male sexual function, and sex reversal should not be considered (see Chapter 14).

Birth injury, asphyxia, or breech delivery is more common in the neonatal history of patients with idiopathic hypopituitarism, especially affected males. While some subjects with breech delivery have central nervous system MRI abnormalities including interrupted pituitary stalk and ectopic posterior pituitary gland, cause and effect is not established as some individuals with breech delivery and hypopituitarism have no such MRI findings.

D. Miscellaneous disorders

1. **Chronic disease and malnutrition**—A delay in sexual maturation may be due to chronic disease or malnutrition. For example, children with intractable asthma have delayed pubertal development leading to short stature during the teenage years, although they ultimately reach an appropriate adult height for family. Children with other chronic diseases may not fare so well in long-term follow-up; for example, HIV infection

in adolescence causes poor growth and poor pubertal progression. Weight loss to less than 80% of ideal body weight, caused by disease or voluntary dieting, may result in gonadotropin deficiency; weight gain toward the ideal usually restores gonadotropin function but there is often a delay after the weight gain.

Chronic disease may have effects on sexual maturation separate from nutritional state. For example, there is a high incidence of hypothalamic hypogonadism in thalassemia major, even with regular transfusion and chelation therapy.

Celiac disease can be found in approximately 1% of the general population and in about 10% of type I diabetes patients. This immunological disorder of gluten intolerance produces numerous clinical sequelae including late puberty, poor growth, delayed menarche, and osteopenia.

2. **Weight loss and anorexia nervosa**—Anorexia nervosa involves weight loss associated with a significant psychologic disorder. This condition usually affects girls who develop a disturbed body image and exhibit typical behavior such as avoidance of food and induction of regurgitation after ingestion. Weight loss may be so severe as to cause fatal complications such as immune dysfunction, fluid and electrolyte imbalance, or circulatory collapse. Primary or secondary amenorrhea is a classic finding in these patients and has been correlated with the degree of weight loss, although there is evidence that patients with anorexia nervosa may cease to menstruate before their substantial weight loss is exhibited. Other endocrine abnormalities in anorexia nervosa include elevated serum GH and decreased IGF-I (these are characteristic of other types of starvation), decreased serum triiodothyronine, decreased serum 1,25-dihydroxyvitamin D and elevated 24,25-hydroxyvitamin D levels. Weight gain to the normal range for height, however, does not ensure immediate resumption of menses. There is an increased incidence of anorexia nervosa in ballet dancers or ballet students; the incidence of scoliosis and mitral valve insufficiency is also increased in these patients. Functional amenorrhea may also occur in women of normal weight, some of whom demonstrate evidence of psychologic stress. Decreased LH response to GnRH administration, impaired monthly cycles of gonadotropin secretion, and retention of a diurnal rhythm of gonadotropin secretion are found in anorexia nervosa patients, patterns that indicate a reversion to an earlier stage of the endocrine changes of puberty.

3. **Increased physical activity**—Girls who regularly participate in activities such as strenuous athletics and ballet dancing may have delayed thelarche, delayed menarche, and irregular or absent menstrual periods. Increased physical activity and not decreased weight appears to be the cause of the amenorrhea in some girls; such amenorrheic patients may resume menses while temporarily bedridden and inactive even though weight does not yet change. Statistical analysis of mothers' ages at menarche and the type of sport pursued by their daughters indicates that late maturation of some gymnasts may be referable to a familial tendency to late menarche, suggesting that gymnastic activity may not be the primary cause of late menarche. On the other hand, there are studies indicating that intensive gymnastics training at an early age leads to a decrease in ultimate stature.

4. **Hypothyroidism**—Hypothyroidism can delay all aspects of growth and maturation, including puberty and menarche. Galactorrhea may occur in severe primary hypothyroidism due to concomitant elevation of serum PRL due to TSH stimulation. With thyroxine therapy, catch-up growth and resumed pubertal development and menses occur. Conversely, severe primary hypothyroidism may be associated with precocious puberty and galactorrhea (due to elevated serum PRL) in some patients (Van Wyk-Grumbach syndrome). There is evidence that excessive TSH can stimulate FSH receptors, leading to estrogen secretion and breast development in these patients.

5. **Genetic syndromes**
   a. **Prader-Willi syndrome**—Prader-Willi syndrome (OMIM#176270) occurs sporadically and is associated with fetal and infantile hypotonia, short stature, poor feeding in infancy but insatiable hunger later, leading to marked obesity, characteristic facies with almond-shaped eyes, small hands and feet after infancy, developmental delay, and emotional instability in patients of either sex. Delayed menarche in females and micropenis and cryptorchism in males is common; while there is hypothalamic hypogonadism, there is also testicular dysfunction reported in males. Osteoporosis is common in these patients during the teenage years, but sex steroid replacement, when indicated, may increase bone density, although the effects are not proven to be long lasting if sex steroid administration ceases. Behavioral modification may improve the usual pattern of rampant weight gain, but the constant vigilance of caregivers is necessary. Patients have absence of activity of paternally derived gene at 15q11-13 due to deletion of the area on the father's chromosome, uniparental disomy (both areas derived from mother; UPD) or defects in the imprinted area. If an abnormality in this area derives from the mother, Angelman syndrome results. Genetic testing starts with a methylation test for 15q11.2-q13, which identifies over 99% of those with the syndrome; genetic consultation is requested for those suspected of having the syndrome but not diagnosed with methylation testing. This test is available commercially for diagnosis.
   b. **Bardet-Biedl syndromes**—This autosomal recessive but genetically heterogeneous condition is characterized by obesity, short stature, developmental delay, kidney dysfunction, polydactyly, and retinitis pigmentosa. Hypogonadotropic hypogonadism and primary hypogonadism have variously been reported in affected patients. There are presently 19 identified types of this condition listed in OMIM.

Hypergonadotropic Hypogonadism

Primary gonadal failure is heralded by elevated gonadotropin concentrations due to the absence of negative feedback effects of gonadal sex steroids and inhibin. The most common causes of hypergonadotropic hypogonadism are associated with chromosomal and somatic abnormalities (see Chapter 14), but isolated gonadal failure can also present with delayed puberty without other physical findings. When hypergonadotropic hypogonadism is present in patients with a Y chromosome or a fragment of a Y chromosome (genetic males or conditions noted later), testicular dysgenesis must be considered in the differential diagnosis. The risk of testicular cancer rises in testicular dysgenesis (testicular cancer in normal boys is rare; eg, the incidence in Scandinavia is rising but still 0.5 per 100,000 in childhood).

A. Syndrome of seminiferous tubule dysgenesis (Klinefelter syndrome)   A common form of primary testicular failure is Klinefelter syndrome (47,XXY karyotype), with an incidence of 1:1000 males. Before puberty, patients with Klinefelter

syndrome have decreased upper segment-lower segment ratios, small testes, and an increased incidence of developmental delay and personality disorders. Onset of puberty is not usually delayed, because Leydig cell function is characteristically less affected than seminiferous tubule function in this condition and testosterone is adequate to stimulate pubertal development. Serum gonadotropin levels rise to castrate concentrations after the onset of puberty; the testes become firm and are rarely larger than 3.5 cm in diameter. After the onset of puberty, there are histologic changes of seminiferous tubule hyalinization and fibrosis, adenomatous changes of the Leydig cells, and impaired spermatogenesis. Gynecomastia is common, and variable degrees of male secondary sexual development are found. Surviving sperm have been found by microdissection of some seminiferous tubules in a few patients leading to successful fertilization of a partner's ovum. Other forms of male hypergonadotropic hypogonadism are found with 46,XX/47,XXY, 48,XXYY, 48,XXXY, and 49,XXXXY karyotypes. Phenotypic males are described with 46,XX karyotypes and some physical features of Klinefelter syndrome; they may have the *SRY* gene translocated to an X chromosome (see Chapter 14).

B. Other forms of primary testicular failure Patients surviving treatment for malignant diseases form a growing category of testicular failure. Chemotherapy—primarily with alkylating agents—or radiation therapy directed to the gonads may lead to gonadal failure. Injury is more likely if treatment is given during puberty than if it occurs in the prepubertal period, but even prepubertal therapy leads to risk. Normal pubertal development may occur in boys treated with chemotherapy before puberty, although they demonstrate elevated peak serum LH and elevated basal and peak serum FSH after GnRH, as well as a high incidence of decreased or absent sperm counts. This indicates that prepubertal status does not protect a child against testicular damage from chemotherapy and that normal physical development may hide significant endocrine and reproductive damage.

The **Sertoli cell only syndrome** (germinal cell aplasia) is a congenital form of testicular failure manifested by azoospermia and elevated FSH concentrations but generally normal secondary sexual characteristics, normal testosterone concentrations, and no other anomalies. A mutation in a gene at Yq11.23, the azoospermia factor *(AZF)* (OMIM#415000 spermatogenic failure, nonobstructive, Y-linked) appears to play a role in the lack of production of spermatocytes.

Patients with Down syndrome may have elevated LH and FSH levels even in the presence of normal testosterone levels, suggesting some element of primary gonadal failure.

Advances in the treatment of cancer allow long-term survival past childhood. Chemotherapy and radiation therapy often affect the gonads and lead to infertility. The field of oncofertility is rapidly advancing and promises the potential for fertility preservation. Sperm banking is possible before therapy in males and there are advances in ovarian preservation for girls.

C. Cryptorchism or anorchia Phenotypic males with a 46,XY karyotype but no palpable testes have either cryptorchism or anorchia. Cryptorchid males should produce a rise in

testosterone levels greater than 2 ng/dL within 72 hours after intramuscular administration of hCG (3000 U/m^2) in the newborn period, although repeated doses are required after the first postnatal months. Testes may also descend during 2 weeks of treatment with hCG given three times a week.

Patients with increased plasma testosterone levels in response to hCG administration, but without testicular descent, have cryptorchism. Their testes should be brought into the scrotum by surgery to decrease the likelihood of further testicular damage due to the elevated intra-abdominal temperature and to guard against the possibility of undetected tumor formation. Cryptorchid testes may demonstrate congenital abnormalities and may not function normally even if brought into the scrotum early in life. Furthermore, the descended testis in a unilaterally cryptorchid boy may itself show abnormal histologic features; such patients have a 69% incidence of decreased sperm counts. Thus, unilateral cryptorchid patients can be infertile even if they received early treatment of their unilateral cryptorchism. Since both the descended and the undescended testes may be affected, there may be preexisting disease that is manifested by the lack of descent of one testis. In addition, patients undergoing orchiopexy may sustain subtle damage to the vas deferens, leading to the later production of antibodies to sperm that may result in infertility. This finding may depend on the surgical technique used.

It is important to determine if any testicular tissue is present in a boy with no palpable testes, because unnoticed malignant degeneration of the tissue is a possibility. The diagnosis of anorchia due to the testicular regression syndrome may be pursued by ultrasound, MRI, laparotomy, or laparoscopic examination. There are other endocrine evaluations besides the hCG test that may help in diagnosis. The presence of anti-Müllerian factor in a young child indicates the presence of testicular tissue, although during puberty anti-Müllerian factor becomes undetectable, and this test cannot be employed at that stage. Inhibin may be measured as an indication of functioning testicular tissue. Except for the absence of testes, 46,XY patients who are otherwise normal and have the *vanishing testes* syndrome of late fetal loss of the testes have normal infantile male genital development, including Wolffian duct formation and Müllerian duct regression. The testes were presumably present in these patients early in fetal life during sexual differentiation, but degenerated after the 13th week of gestation for unknown reasons (see Chapter 14). When these boys reach adulthood, they are reported to establish a normal male gender identity. The presence of normal basal gonadotropin levels in a prepubertal boy without palpable testes at least suggests the presence of testicular tissue even if the testosterone response to hCG is low, whereas the presence of elevated gonadotropin levels without any testosterone response to hCG suggests anorchia.

The rate of cryptorchidism has increased from 2.7% in the 1950s to 5% in the early 21st century in England and Denmark with a corresponding rise in testicular carcinoma. Likewise an international trend to lower sperm production and decreased testosterone values appear associated with a falling fertility rate found in Europe and Japan. The hypothesis of the testicular dysgenesis syndrome (TDS) postulates that environmental exposures to

endocrine disruptors in utero and later lead to these significant changes in reproductive function.

D. Syndrome of gonadal dysgenesis (Turner syndrome)

45,XO gonadal dysgenesis is associated with short stature, female phenotype with sexual infantilism. Patients have streak gonads consisting of fibrous tissue without germ cells. Other classic but variable phenotypic features include micrognathia, *fish* mouth (downturned corners of the mouth), ptosis, low-set or deformed ears, a broad shield-like chest with the appearance of widely spaced nipples, hypoplastic areolae, a short neck with low hairline and webbing of the neck (pterygium colli), short fourth metacarpals, cubitus valgus, structural anomalies of the kidney, extensive nevi, hypoplastic nails, and vascular anomalies of the left side of the heart (most commonly coarctation of the aorta associated with hypertension proximal to the coarctation). However, some affected girls can have almost normal female phenotypes. The medical history of patients with the syndrome of gonadal dysgenesis often reveals small size at birth, lymphedema of the extremities most prominent in the newborn period, and loose posterior cervical skin folds. (The terms *Bonnevie-Ullrich syndrome* and *infant Turner syndrome* are applied to this neonatal appearance.) Affected patients often have a history of frequent otitis media with conductive hearing loss. There may be frequent urinary tract infections associated with a horseshoe-shaped kidney, duplication of ureters, or other anatomic defects. Intelligence is normal, but there is often impaired spatial orientation, which may lead to difficulty in mathematics, especially geometry. Patients have no pubertal growth spurt, and reach a mean adult height of 143 cm. Short stature is a classic feature of Turner syndrome, but not of other forms of hypergonadotropic hypogonadism that occur without karyotype abnormalities. The short stature is linked to the absence of the (*short stature homeobox*) *SHOX* homeobox gene of the pseudoautosomal region of the X chromosome (Xpter-p22.32) (OMIM*312865). GH function is usually normal in Turner syndrome. However, exogenous hGH treatment improves growth rate and increases adult stature toward the normal range in affected girls. It improves lipid profiles and decreases diastolic blood pressure (see Chapter 6). Pubic hair may appear late and is usually sparse in distribution owing to the absence of any ovarian secretions; adrenarche progresses in Turner syndrome even in the absence of gonadarche. Autoimmune thyroid disease (usually hypothyroidism) is common in Turner syndrome, and determination of thyroid function and thyroid antibody levels is important in evaluation of these patients.

Serum gonadotropin concentrations in Turner syndrome are extremely high between birth and about age 4 years. They decrease toward the normal range in prepubertal patients in the juvenile pause and then rise again to castrate levels after age 10 years (see Chapter 14).

Sex chromatin–positive variants of gonadal dysgenesis include 45,X/46,XX, 45,X/47,XXX, and 45,X/46,XX/47,XXX mosaicism with chromatin-positive buccal smears. Patients with these karyotypes may resemble patients with the classic syndrome of gonadal dysgenesis, or they may have fewer manifestations and normal or nearly normal female phenotypes. Streak gonad formation is not invariable; some patients have had secondary sexual development, and menarche and rare pregnancies have been reported.

Sex chromatin–negative variants of the syndrome of gonadal dysgenesis have karyotypes with 45,X/46,XY mosaicism. Physical features vary; some patients have the features of classic Turner syndrome, whereas others may have ambiguous genitalia or even the features of phenotypic males. Gonads are usually dysgenetic but vary from streak gonads to functioning testes. These patients are at risk for gonadoblastoma formation; this is a benign tumor that has the potential for malignant transformation and then may metastasize. Because gonadoblastomas may secrete androgens or estrogens, patients with gonadoblastoma may virilize or feminize as though they had functioning gonads, confusing the clinical picture. Gonadoblastomas may demonstrate calcification on abdominal x-ray. Malignant germ cell tumors may arise in dysgenetic testes, and gonadectomy (orchiectomy is a more precise term) is generally indicated. In some mosaic patients with one intact X chromosome and one chromosomal fragment, it is difficult to determine whether the fragment is derived from an X chromosome or a Y chromosome. Polymerase chain reaction techniques to search for specific Y chromosome sequences are helpful if a karyotype reveals no Y chromosomal material.

Patients with Turner syndrome who desire fertility can be treated with *in vitro* fertilization techniques. After exogenous hormonal preparation, a fertilized ovum (possibly a sister's ovum fertilized by the patient's male partner, or an extra fertilized ovum from another couple undergoing *in vitro* fertilization) can be introduced into the patient's uterus, and the pregnancy is then brought to term by support with exogenous hormone administration. A serious possible complication is uterine rupture which makes this a very dangerous procedure.

E. Other forms of primary ovarian failure

Ovaries appear to be more resistant to damage from the chemotherapy used in the treatment of malignant disease than are testes. Nonetheless, ovarian failure can occur with medical therapy. Damage is common if the ovaries are not surgically moved out of the path of the beam or shielded by lead in abdominal radiation therapy. Normal gonadal function after chemotherapy does not guarantee normal function later. Late-onset gonadal failure has been described after chemotherapy. Bone marrow transplantation with whole-body irradiation for acute lymphoblastic leukemia or non–Hodgkin lymphoma appears to cause the most damage to endocrine function.

Premature menopause has also been described in otherwise healthy girls owing to the presence of anti-ovarian antibodies. Patients with Addison disease may have autoimmune oophoritis as well as adrenal failure. A sex steroid biosynthetic defect due to 17α-hydroxylase deficiency (P450c17) manifests as sexual infantilism and primary amenorrhea in a phenotypic female (regardless of genotype) with hypokalemia and hypertension (see Chapter 14). A patient with 17α-hydroxylase deficiency may have ovaries or testes and still present as a phenotypic female. The syndrome of androgen resistance may present as a female phenotype with amenorrhea, in a manner suggesting ovarian failure, but the problem in such an individual with a 46,XY karyotype is the absence of androgen receptors.

F. Noonan syndrome (pseudo–Turner syndrome, Ullrich syndrome, male Turner syndrome [OMIM#163950 Noonan syndrome 1; NS1]) Noonan syndrome is associated with some phenotypic manifestations of Turner syndrome such as webbed neck, ptosis, short stature, cubitus valgus, and lymphedema, but other additional clinical findings such as a normal karyotype, triangular facies, pectus excavatum, right-sided heart disease, and an increased incidence of developmental delay differentiate these patients from those with Turner syndrome. Males may have undescended testes and variable degrees of germinal cell and Leydig cell dysfunction. Noonan syndrome follows an autosomal dominant pattern of inheritance with incomplete penetrance (gene locus 12q24). hGH is approved by the FDA to increase stature in these patients.

G. Familial and sporadic forms of 46,XX or 46,XY gonadal dysgenesis These forms of gonadal dysgenesis are characterized by structurally normal chromosomes and streak gonads or partially functioning gonads. They do not have the physical features of Turner syndrome. If there is some gonadal function, 46,XY gonadal dysgenesis may present with ambiguous genitalia or virilization at puberty; 46,XY complete gonadal dysgenesis is also known as Swyer Syndrome. If no gonadal function is present, patients appear as phenotypic sexually infantile females. Patients with 46,XY gonadal dysgenesis and dysgenetic testes should undergo gonadectomy to eliminate the possibility of malignant germ cell tumor formation.

H. Primary amenorrhea associated with normal secondary sexual development If a structural anomaly of the uterus or vagina interferes with the onset of menses but the endocrine milieu remains normal, the patient presents with primary amenorrhea in the presence of normal breast and pubic hair development. A transverse vaginal septum may seal the uterine cavity from the vaginal orifice, leading to the retention of menstrual flow—as may an imperforate hymen. The Rokitansky-Küster-Hauser syndrome (OMIM#277000) combines congenital absence of the vagina with abnormal development of the uterus, ranging from a rudimentary bicornuate uterus that may not open into the

vaginal canal to a virtually normal uterus; surgical repair may be possible in patients with minimal anatomic abnormalities, and fertility has been reported. Associated abnormalities include major urinary tract anomalies and spinal or other skeletal disorders. The rarest anatomic abnormality in this group is absence of the uterine cervix in the presence of a functional uterus.

46,XY disorder of sexual development (DSD), previously called male pseudohermaphroditism, is an alternative cause of primary amenorrhea if a patient has achieved thelarche. The syndrome of complete androgen resistance leads to female external genitalia and phenotype without axillary or pubic hair development in the presence of pubertal breast development (syndrome of testicular feminization; see Chapter 14).

Differential Diagnosis of Delayed Puberty (Table 15–3)

Patients who do not begin secondary sexual development by age 13 (girls) or age 13.5–14 (boys) and patients who do not progress through development on a timely basis (girls should menstruate within 5 years after breast budding; boys should reach stage 5 pubertal development 4 and a half years after onset) should be evaluated for hypogonadism. The yield of diagnosable conditions is quite low in children younger than these ages, but many patients and families request evaluation well before these limits. Without significant signs or symptoms of disorders discussed above, it is best to resist evaluation and offer support until these ages in most cases.

A patient with constitutional delay may have a characteristic history of short stature for age with normal growth velocity for bone age and a family history of delayed but spontaneous puberty. The patient's mother may have had late onset of menses, or the father may have begun to shave late or continued growing after high school. Not all patients with constitutional delay are so classic, and gonadotropin-deficient patients may have some features similar to those of constitutional delay in adolescence. Indeed, patients with Kallmann syndrome and others with constitutional delay are occasionally found in the same kindred.

If the diagnosis is not obvious on the basis of physical or historical features, the differential diagnostic process begins with

TABLE 15–3 Differential diagnosis of delayed puberty.

	Serum Gonadotropins	Serum Gonadal Steroids[a]	Miscellaneous
Constitutional delay in growth and adolescence	Prepubertal (low)	Low	Patient usually has short stature for chronologic age but appropriate height and growth rate for bone age. Both adrenarche and gonadarche are delayed.
Hypogonadotropic hypogonadism	Prepubertal (low)	Low	Patient may have anosmia (Kallmann syndrome) or other associated pituitary hormone deficiencies. If gonadotropin deficiency is isolated, the patient usually has normal height and growth rate but lacks a pubertal growth spurt. Adrenarche may occur at a normal stage in spite of absent gonadarche (serum DHEAS may be pubertal).
Hypergonadotropic hypogonadism	Elevated	Low	Patient may have abnormal karyotype and stigmatas of Turner syndrome or Klinefelter syndrome.

[a]In a pediatric specific assay.

determination of whether plasma gonadotropins are (1) elevated due to primary gonadal failure or (2) decreased due to secondary or tertiary hypogonadism or constitutional delayed puberty. Gonadotropins must be measured in a third-generation assay with pediatric standards, as the customary gonadotropin determinations are too insensitive to reveal the small changes of puberty. A single ultrasensitive, third-generation determination of serum LH concentration in the pubertal range suggests that puberty is progressing. Determination of the rise in LH after administration of GnRH or GnRH agonist is helpful in differential diagnosis; secondary sexual development usually follows within months after conversion to a pubertal LH response to GnRH. Because it is difficult, if not impossible, to obtain GnRH presently, GnRH agonists are frequently used for diagnosis with serum LH values measured after administration.

Values are laboratory-specific, but it has been reported that a basal LH value over 0.3 U/L or a peak LH over 3.2 U/L one hour after 10 μg/kg subcutaneous GnRH administration indicates the onset of puberty in girls with 90% specificity with a stimulated LH over 5.5 U/L providing 96% specificity. A 4-hour LH value over 14.8 U/L after 10 μg/kg of leuprolide acetate in boys provides 100% sensitivity, while a value over 19 U/L provides 100% specificity in identifying the onset of puberty.

Frequent nighttime sampling (every 20 minutes through an indwelling catheter) to determine the amplitude of peaks of LH secretion during sleep is an alternative to GnRH or GnRH agonist testing but is too cumbersome for a clinical settings. Unfortunately, the results of GnRH infusions or nighttime sampling are not always straightforward. Patients may have pubertal responses to exogenous GnRH but may not spontaneously secrete adequate gonadotropins to allow secondary sexual development. In females with amenorrhea, the frequency and amplitude of gonadotropin secretion may not change to allow monthly menstrual cycles. The retention of a diurnal rhythm of gonadotropin secretion (normal in early puberty) into late puberty is a pattern linked to inadequate pubertal progression. In males, a morning serum testosterone concentration over 20 ng/dL (0.7 mmol/L) indicates the likelihood of pubertal development within 6 months. Measurement of testosterone or estradiol must be done in an ultrasenstive assay with pediatric standards or, as with gonadotropin assays, the small changes of pubertal development will be missed. Presently highly sensitive high performance liquid chromatography (HPLC) tandem mass spectroscopy (HPLC MS/MS) assays are available at national laboratories for this purpose.

Other methods of differential diagnosis between constitutional delay and hypogonadotropic hypogonadism have been proposed but are complex or are not definitive. Clinical observation for signs of pubertal development and laboratory evaluation for the onset of rising levels of sex steroids may have to continue until the patient is 18 years of age before the diagnosis is definite. In most cases, if spontaneous pubertal development is not noted by 18 years of age, the diagnosis is gonadotropin deficiency. Of course, the presence of neurologic impairment or other endocrine deficiencies should immediately lead to investigation for central nervous system tumor or congenital defect in a patient with delayed puberty. MRI scanning is invoked in this situation.

Treatment of Delayed Puberty

A. Constitutional delay in growth and adolescence

Psychologic support   Teenagers who are so embarrassed about short stature and lack of secondary sexual development that they have significant psychologic problems may require psychological evaluation and therapy if they have passed the ages of 13 years for girls or 14 years for boys. Patients with constitutional delay in growth and adolescence should be counseled that normal pubertal development will occur spontaneously. Peer pressure and teasing can be oppressive. Although the majority of these patients do quite well, severe depression must be treated appropriately, because short patients with pubertal delay have become suicidal. In some cases, it helps to excuse the patient from physical education class, because the lack of development is most apparent in the locker room. In general, boys feel more stress than girls when puberty is delayed.

Sex steroids   New preparations of topical sex steroids, that have different systemic effects and length of action than parenteral or oral sex steroids, have enlarged the therapeutic armamentarium for delayed puberty. The classic recommendation for girls has been a 3-month course of conjugated estrogen (0.3 mg) or ethinyl estradiol (5-10 μg) given orally each day. Topical estrogen is said to carry less risk of hypertension, gallstones, increased fat mass, decreased insulin sensitivity, and increased triglycerides (topical estrogen does not, however, increase HDL cholesterol and decrease low-density lipoprotein [LDL] cholesterol, as does oral estrogen). Estrogen-impregnated patches are available and might be used to initiate puberty, but the lowest dose available leads to serum estrogen values higher than physiologic levels found in early puberty. They might be used for only a few days each week to reduce the delivered dose. Alternatively, 17-beta estradiol patches (0.025 mg/patch; Vivelle Dot matrix patch) may be cut into fragments the size of one-eighth to one-fourth of a patch fragments to initiate pubertal development with more physiologic estrogen values. However, the use of such topical estrogen preparations is not approved by the Food and Drug Administration (FDA) for induction of puberty.

The classic recommendation for boys is a 3-month course of testosterone enanthate or cypionate (50 mg) given intramuscularly once every 28 days for three doses. The possibility that a higher dose will cause priapism limits the initial dose to 50 mg. The initial doses are increased over months to 100 or 150 mg every 28 days or lower doses every 14 days. Patch or testosterone gel preparations could be used to initiate puberty, but the dose in manufactured packets or pumps is too high to mimic the physiologic levels found in early puberty as these preparations are made for full replacement in adult males. The entire gel packet or lowest-dose patch could be used every other day over the 3-month period of treatment, although this leads to variable daily dosage. The gel available in a pump may be approximated to a lower dose of one-half a single pump depression. The testosterone patches are not impregnated like the estrogen patches and cannot be cut down in size. Thus, to date, the low-dosage injection of testosterone or the use of an estimated one-half or less of the 5 mg

testosterone gel packet or pump are the only methods of reaching appropriate dosage. The topical use of testosterone is not approved by the FDA for pubertal onset. Therapy with oxandrolone has been suggested as a method of increasing secondary sexual development and increasing growth without advancing skeletal development. This method is not widely preferred. Furthermore, testosterone, which can be aromatized to estradiol, increases the generally low endogenous GH secretion in constitutional delayed puberty to normal, whereas oxandrolone, which cannot be aromatized, does not increase GH secretion.

Hormonal treatment of boys or girls elicits noticeable secondary sexual development and a slight temporary increase in stature. The low doses recommended do not advance bone age substantially and do not significantly change adult height if used for 3 months. Low-dose sex steroid treatment has been reported to promote spontaneous pubertal development after sex steroid therapy is discontinued, although responding individuals may include those boys who are on the brink of further pubertal development and are, therefore, most likely to achieve a growth response to androgen therapy. However, a short course of therapy may improve patients' psychologic outlook and allow them to await spontaneous pubertal development with greater confidence. Continuous gonadal steroid replacement in these patients is not indicated, because it advances bone age and leads to epiphyseal fusion and a decrease in ultimate stature. After a 3 to 6 month break to observe spontaneous development, a second course of therapy may be offered if no spontaneous pubertal progression occurs during observation.

B. Persistent hypogonadism Once a patient has been diagnosed as having delayed puberty due to apparently permanent primary or secondary hypogonadism, replacement therapy must be considered at the average age of onset of puberty in most cases.

Males with hypogonadism may be treated with testosterone gel, testosterone patches, or testosterone enanthate or cypionate given intramuscularly every month, as described earlier for constitutional delay. Testosterone gel or patches, as described earlier, are used for persistent, long-term replacement but are not FDA-approved for individuals less than 16 years. Treatment as for constitutional delay should be started and gradually increased to the adult range over months to years to mimic the normal progression of puberty and to avoid abrupt exposure to high-dose androgen and the possibility of frequent erections or priapism. Oral halogenated testosterone and methylated testosterone are never recommended because of the risk of hepatocellular carcinoma or cholestatic jaundice. Testosterone therapy may not cause adequate pubic hair development, but patients with secondary or tertiary hypogonadism may benefit from hCG administration with increased pubic hair growth resulting from endogenous testicular androgen secretion in addition to the exogenous testosterone.

Girls may be treated with oral ethinyl estradiol (increasing from 5 µg/d to 10 to 20 µg/d depending on clinical results), or conjugated estrogens (0.3 or 0.625 mg/d). One may start with daily doses with eventual conversion to treatment on days 1 to 21 of the month after several months of daily ingestion when vaginal secretions increase. Five to 10 mg of medroxyprogesterone acetate

are then added on days 12 to 21 after physical signs of estrogen effect are noted and breakthrough bleeding occurs (and always within 6 months after initiating estrogen). Neither hormone is administered from day 22 to the end of the month to allow regular withdrawal bleeding (see Chapter 13). Later, the patient may be switched to sequential oral contraceptives to simplify treatment. As described earlier, estrogen patch therapy may have benefits (but these are not proven) over oral estrogen and may be preferable (but these are not FDA-approved for individuals under 16 years of age). Low-dose estrogen oral contraceptives may be used after initial induction phase. Gynecologic examinations should be performed yearly for those on replacement therapy.

Hypothalamic hypogonadism may be treated with GnRH pulses by programmable pumps to achieve fertility or to promote pubertal development. Likewise, in the absence of a functional pituitary gland, hCG and menotropins (human postmenopausal gonadotropin) may be administered in pulses. These techniques are cumbersome and best reserved for the time when fertility is desired.

C. Coexisting GH deficiency The treatment of patients with coexisting GH deficiency requires consideration of their bone age and amount of growth left before epiphyseal fusion. If such a patient has not yet received adequate treatment with GH, sex steroid therapy may be kept at the lower dosage or even delayed to optimize adult height. The goal is to allow appropriate pubertal changes to support psychologic development and to allow the synergistic effects of combined sex steroids and GH without fusing the epiphyses prematurely.

Constitutional delayed puberty may be associated with decreased GH secretion in 24-hour profiles of spontaneous secretion or in stimulated testing. GH secretion increases when pubertal gonadal steroid secretion rises due to the effects of estrogen, so decreased GH secretion in this condition should be considered temporary. Nonetheless, true GH-deficient patients may have delayed puberty due to the GH deficiency or to coexisting gonadotropin deficiency. Therefore, deciding whether a pubertal patient has temporary or a permanent GH deficiency can be difficult; previous observation may indicate a long history characteristic of constitutional delay in adolescence, whereas a recent decrease in growth rate may suggest the onset of a brain tumor or other cause of hypopituitarism.

D. The syndrome of gonadal dysgenesis In the past, patients with the syndrome of gonadal dysgenesis were frequently not given estrogen replacement until after age 13 years, for fear of compromising adult height. However, low-dose estrogen therapy (5-10 µg of ethinyl estradiol orally) can be administered to allow feminization and improve psychologic status at 12 to 13 years of age without decreasing final height, as shown in several studies. Low-dose estrogen increases growth velocity, whereas high-dose estrogen suppresses it. Even if growth velocity is increased, however, adult height is not increased with such estrogen treatment. Treatment of Turner syndrome with GH successfully increases adult stature (see Chapters 6 and 14). During childhood these girls must be regularly screened for strabismus, hearing loss, and

autoimmune thyroid disease as well as learning difficulties. There are several recommendations for the preparation of these girls to transition to adult care and necessary follow-up for issues that will affect their later life, including monitoring for the development of aortic dilation at 5- to 10-year intervals since this defect may progress to aortic aneurysms.

E. Bone mass   After infancy, most bone mass accrual takes place during the second decade of life, and disorders of puberty may affect this process. Delayed puberty in boys causes decreased cross-sectional bone density, when the subjects are tested as young adults, because they have not yet reached peak bone acquisition. Although there is some controversy, it appears likely that normalization of volumetric bone density occurs with maturity and exposure to a normal androgen milieu. A range of defects in girls such as anorexia nervosa, athletics-induced delayed puberty, and Turner syndrome also cause decreased bone density. The use of testosterone in boys and estrogen and progesterone in girls with hypogonadotropic hypogonadism is helpful in increasing bone mass but has not been demonstrated to result in normal adult bone mass. Appropriate ingestion of dairy products containing calcium or calcium supplementation and vitamin D should be encouraged in hypogonadal or constitutionally delayed patients as well as in normal children at least as a common sense measure. No long-term follow-up is yet available, however, to prove the efficacy of this therapy in susceptible subjects.

PRECOCIOUS PUBERTY (SEXUAL PRECOCITY)

All sources agree that the appearance of secondary sexual development before the age of 9 years in boys is precocious puberty. However, there remains controversy over the lower limits of normal in girls. The Pediatric Endocrine Society states that the appearance of secondary sexual development before the age of 7 years in Caucasian girls and 6 years in African American girls constitutes precocious sexual development but others remain concerned that pubertal development between 7 and 8 years for Caucasian and 6 and 8 for African American girls indicate a pathological state (Table 15–4). However, if a girl has the onset of puberty before 8 years, one must have a high index of suspicion for pathology. The child must have absolutely no sign of central nervous system disorder or other possible cause that might trigger pathologic precocious puberty. A careful search for historical or physical features of organic disease must occur before a girl with precocious puberty before 8 years is considered normal. When the cause of precocious puberty is premature activation of the hypothalamic-pituitary axis and the condition is gonadotropin-dependent, the diagnosis is **central (complete or true) precocious puberty**; if ectopic gonadotropin secretion occurs in boys or autonomous sex steroid secretion occurs in either sex, the condition is not gonadotropin-dependent, and the diagnosis is peripheral or **incomplete precocious puberty**. If feminization occurs in girls and virilization occurs in boys, the condition is **isosexual**

TABLE 15–4 Classification of precocious puberty.

Central (complete or true) GnRH-dependent isosexual precocious puberty
Constitutional
Idiopathic
Central nervous system disorders (including congenital defects)
Tumors
Infection
Trauma
Radiation
Following androgen exposure
Peripheral or incomplete GnRH-independent isosexual precocious puberty
Males
Gonadotropin-secreting tumors
Excessive androgen production
Testicular or adrenal tumors
Virilizing congenital adrenal hyperplasia
Premature Leydig and germinal cell maturation
Females
Ovarian cysts
Estrogen-secreting neoplasms
Severe hypothyroidism
Males and females
McCune-Albright syndrome
Peripheral or incomplete contrasexual precocity
Males
Estrogen-secreting tumor
Females
Androgen-secreting tumor
Virilizing congenital adrenal hyperplasia
Iatrogenic sexual precocity due to gonadotropin or sex steroid exposure
Variation in pubertal development
Premature thelarche
Premature menarche
Premature pubarche
Adolescent gynecomastia

precocity, but if the feminization occurs in boys and virilization in girls, the condition is **contrasexual precocity**. In all forms of sexual precocity, there is an increase in growth velocity, somatic development, and skeletal maturation. When unchecked, this rapid skeletal development may lead to tall stature during the early phases of the disorder but to short final stature because of early epiphysial fusion. This is the paradox of the tall child growing up to become a short adult. Plasma IGF-I values may be elevated for age but are appropriate for pubertal stage in untreated precocious puberty.

Central (Complete or True) Precocious Puberty (Figure 15–6)

A. Constitutional or familial central (complete or true) precocious puberty   Otherwise normal children who demonstrate isosexual precocity at an age slightly more than 2.5 SD below the mean may simply represent the lower reaches of the distribution curve describing the age of onset of puberty; often there is a familial tendency toward early puberty. True precocious
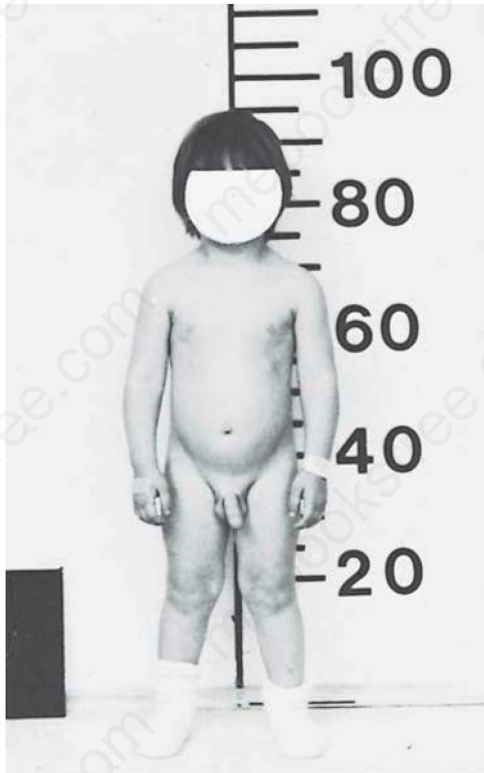
FIGURE 15–6 Boy 2-5/12 years of age with idiopathic true precocious puberty. By 10 months of age, he had pubic hair and phallic and testicular enlargement. At 1 year of age, his height was 4 SD above the mean; the phallus was 10 cm × 3.5 cm; each testis was 2.5 cm × 1.5 cm. Plasma LH was pubertal and rose in an adult pattern after administration of 100 μg of GnRH. Plasma testosterone was 416 ng/dL. At the time of the photograph, he had been treated with medroxyprogesterone acetate (to suppress LH and FSH secretion) for 1½ years, with reduction of his rapid growth rate and decreased gonadotropin and testosterone secretion. At that point his height was 95.2 cm (>2 SD above mean height for his age); plasma testosterone was 7 ng/dL, and after 100 μg of GnRH, plasma LH rose only slightly, demonstrating suppression. (Reproduced with permission from Yen SSC, Jaffe RB. *Reproductive Endocrinology*. Philadelphia: WB Saunders; 1978.)

puberty is reported, rarely, to be due to an autosomal dominant or (in males) X-linked dominant trait.

B. Idiopathic central (complete or true) isosexual precocious puberty Affected children, with no familial tendency toward early development and no organic disease, may be considered to have idiopathic central isosexual precocious puberty. Epilepsy and developmental delay are associated with central precocious puberty in the absence of a central nervous system anatomic abnormality, indicating a central nervous system process is responsible for the condition.

Pubertal development may follow the normal course or may wax and wane. Serum gonadotropin and sex steroid concentrations and response to GnRH or GnRH agonists are similar to those found in normal pubertal subjects. In idiopathic central precocious puberty, as in all forms of true isosexual precocity, testicular enlargement in boys should be the first sign; in girls, breast development or, rarely, pubic hair appearance may be first. Girls present with idiopathic central precocious puberty more commonly than boys. Children with precocious puberty have a tendency toward obesity based on elevated body mass index in the untreated state in some studies.

C. Genetic causes of precocious puberty While families with inactivating mutations of the *GRP54* gene have absent or delayed puberty, a patient presented with isosexual precocious puberty and demonstrated a mutation of *GRP54* that led to prolonged activation of intracellular signaling pathways in response to kisspeptin. This further emphasizes the importance of the GRP54-kisspepetin axis in the control of puberty. A boy is described with a mutation in the gene for *KISS1* leading to prolonged activity of the ligand and precocious puberty.

The *MRKN3* gene is maternally imprinted leading to silencing of the maternal allele. This is considered to exert a braking effect on pubertal development, and thus, when a paternal inactivating mutation occurs, since the maternal allele is silenced, the child experiences precocious puberty. Numerous families are now reported with such mutations with affected girls experiencing a median age of onset of puberty of 6 years while boys are less affected and experience a later onset of puberty.

Other syndromes with chromosomal abnormalities associated with precocious puberty include the Williams-Beuren syndrome, Russell-Silver syndrome, Temple syndrome, Prader Willi syndrome, and reminiscent of Rett syndrome.

D. Central nervous system disorders

1. **Tumors**—Central nervous system tumors are more common causes of central precocious puberty in boys than in girls but may occur in either, so a high index of suspicion is essential. Optic gliomas or hypothalamic gliomas, astrocytomas, ependymomas, germinomas, and other central nervous system tumors may cause precocious puberty by interfering with neural pathways that inhibit GnRH secretion, thus releasing the central nervous system restraint of gonadotropin secretion. Patients with optic gliomas and neurofibromatosis type 1 may have precocious puberty, but no patients with isolated neurofibromatosis (that did not have optic gliomas) are reported with precocious puberty. Remarkably, craniopharyngiomas, which are known to cause delayed puberty, can also trigger precocious pubertal development. Radiation therapy is often indicated in radiosensitive tumors such as germinomas and craniopharyngiomas, where complete surgical extirpation is impossible. Reportedly, complete removal of craniopharyngiomas is associated with unexpected recurrence; radiation therapy appears to be a key therapy for this tumor in addition to surgery.

Hamartomas of the tuber cinereum contain GnRH and neurosecretory cells similar to those found in the median eminence; they may cause precocious puberty by secreting GnRH. Some hypothalamic hamartomas associated with central precocious puberty do not elaborate GnRH but instead contain transforming growth factor (TGF)-α, which stimulates GnRH secretion itself. With improved methods of imaging the central nervous system, hamartomas, with their characteristic radiographic appearance, are now more frequently diagnosed in

patients who were previously thought to have idiopathic precocious puberty. These tumors do not enlarge and so pose no additional threat to the patients in the absence of intractable seizures. These seizures are rare in patients with the pedunculated hamartomas causing central precocious puberty. Because of the location of the hamartoma, surgery is a dangerous alternative to GnRH analog therapy.

Tumors or other abnormalities of the central nervous system may cause GH deficiency in association with central precocious puberty. GH deficiency may also occur after irradiation therapy for such tumors, even if precocious puberty was the sole endocrine finding before irradiation. Such patients grow much faster than isolated GH-deficient patients but slower than children with classic precocious puberty and adequate GH secretion. Often, the GH deficiency is unmasked after successful treatment of precocious puberty. This combination must be considered during the diagnostic process (see also Chapter 6).

2. Other causes of true precocious puberty—Infectious or granulomatous conditions such as encephalitis, brain abscess, postinfectious (or postsurgical or congenital) suprasellar cysts, sarcoidosis, and tuberculous granulomas of the hypothalamus can cause central precocious puberty. Suprasellar cysts and hydrocephalus cause central precocious puberty that is particularly amenable to surgical correction. Brain trauma may be followed by either precocious or delayed puberty. Radiation therapy for acute lymphoblastic leukemia of the central nervous system, or prior to bone marrow transplantation, is characteristically associated with hormonal deficiency, but there are increasing reports of precocious puberty occurring after such therapy. Higher doses of radiation may be more likely to cause GnRH deficiency, and lower doses down to 18 Gy may lead to central precocious puberty.

D. Virilizing syndromes   Patients with long-untreated virilizing adrenal hyperplasia who have advanced bone ages may manifest precocious puberty after treatment with glucocorticoid suppression. Children with virilizing tumors or those given long-term androgen therapy may follow the same pattern when the androgen source is removed. Advanced maturation of the hypothalamic-pituitary-gonadal axis appears to occur with any condition causing excessive androgen secretion and advanced skeletal age.

Peripheral or Incomplete Isosexual Precocious Puberty in Boys

Males may manifest premature sexual development in the absence of hypothalamic-pituitary maturation from either of two causes: (1) ectopic or autonomous endogenous secretion of hCG or LH or iatrogenic administration of hCG, which can stimulate Leydig cell production of testosterone; or (2) autonomous endogenous secretion of androgens from the testes or adrenal glands or from iatrogenic exogenous administration of androgens. (In females, secretion of hCG does not by itself cause secondary sexual development.)

A. Gonadotropin-secreting tumors   These include hepatomas or hepatoblastomas of the liver as well as teratomas or choriocarcinoma of the mediastinum, gonads, retroperitoneum, or pineal gland and germinomas of the central nervous system. The testes are definitely enlarged but not to the degree found in central precocious puberty since only Leydig cells are stimulated by hCG.

B. Autonomous androgen secretion   Secretion of androgens can occur because of inborn errors of adrenal enzyme function, as in 21-hydroxylase deficiency (P450c21 OMIM#201910 adrenal hyperplasia, congenital, caused by 21-hydroxylase) or 11β-hydroxylase deficiency (P450c11β OMIM#202010; adrenal hyperplasia, congenital, caused by steroid 11-beta-hydroxylase deficiency), virilizing adrenal carcinomas, interstitial cell tumors of the testes, or premature Leydig and germinal cell maturation. Late-onset congenital adrenal hyperplasia, generally of the 21-hydroxylase deficiency form, may occur years after birth with no congenital or neonatal manifestations of virilization. Adrenal rest tissue may be found in the testes as a vestige of the common embryonic origin of the adrenal glands and the testes. Congenital adrenal hyperplasia associated with testicular adrenal rest tissue (TART) can enlarge (sometimes to remarkable size) and secrete adrenal androgens (see Chapter 14). TART has been reported to lead to testicular neoplasia.

In boys with familial gonadotropin-independent premature Leydig and germinal cell maturation (OMIM#176410. precocious puberty, male-limited), plasma testosterone levels are in the pubertal range, but plasma gonadotropin levels and the LH response to exogenous GnRH are in the prepubertal range or lower because autonomous testosterone secretion suppresses endogenous GnRH release. The cause of this sex-limited dominant condition lies in a constitutive activation of the LH receptor (leaving the LH receptor on), causing increased cyclic adenosine monophosphate (cAMP) production in the absence of LH and continuous testosterone secretion. Several mutations have been reported in the LH receptor gene in different families (eg, Asp[578] to Gly or Met[571] to Ile). The differential diagnosis rests between testosterone-secreting tumor of the adrenal, testosterone-secreting Leydig cell neoplasm, and premature Leydig and germinal cell maturation.

If hCG secretion due to a tumor causes peripheral male isosexual precocity, FSH is not elevated. As only the Leydig cells are stimulated and because the seminiferous tubules are not stimulated, the testes do not enlarge as much as in complete sexual precocity. If peripheral sexual precocity is due to a testicular tumor, the testes may be large, asymmetric, and irregular in contour. Symmetric bilateral moderate enlargement of the testes suggests familial gonadotropin-independent premature maturation of Leydig and germinal cells, which is a sex-limited dominant condition. The testes are somewhat smaller in this condition, than in true precocious puberty, but are still over 2.5 cm in diameter.

Peripheral or Incomplete Contrasexual Precocity in Boys

Estrogen may be secreted by rare adrenal tumors leading to feminization in boys. Sertoli cell tumors associated with Peutz-Jeghers syndrome is another possible etiology.

Peripheral or Incomplete Isosexual Precocious Puberty in Girls

Females with peripheral isosexual precocity have a source of excessive estrogens. In all cases of autonomous endogenous estrogen secretion or exogenous estrogen administration, serum LH and FSH levels are low.

A. Follicular cysts If follicular cysts are large enough, they can secrete sufficient estrogen to cause breast development and even vaginal withdrawal bleeding; some girls have recurrent cysts that lead to several episodes of vaginal bleeding. Patients with cysts may have levels of serum estrogen high enough to mimic a tumor. Larger follicular cysts can twist on their pedicles and become infarcted, causing symptoms of acute abdomen in addition to the precocious estrogen effects.

B. Other ovarian tumors Juvenile granulosa cell tumor of the ovary is a rare sex chord tumor that can secrete estrogen and lead to precocioius puberty, as 5% are prepubertal. They are palpable in 80% of cases. Gonadoblastomas are found in streak gonads, lipoid tumors, and cystadenomas. Granulosa-theca cell ovarian tumors are even rarer; they secrete androstenedione which can be converted into estrogen. Ovarian carcinomas are rare ovarian sources of estrogens or androgens. Secretory tumors of the ovary may present with abdominal pain, adnexal mass and precocious puberty.

C. Exogenous estrogen administration Ingestion of estrogen-containing substances (eg. oral contraceptives) or even cutaneous absorption of estrogen can cause feminization in children. Epidemics of gynecomastia and precocious thelarche in Puerto Rico and Italy have variously been attributed to ingestion of estrogen-contaminated food, estrogens in the environment, soy formula, or undetermined causes. One outbreak of gynecomastia in boys and precocious thelarche in girls in Bahrain was traced to ingestion of milk from a cow which was given continuous estrogen treatment by its owner to ensure uninterrupted milk production. Tea tree and lavender oils have estrogenic effects in children, and exposure can cause thelarche.

D. Hypothyroidism Severe untreated hypothyroidism can be associated with sexual precocity and galactorrhea (Van Wyk-Grumbach syndrome); treatment with thyroxine corrects hypothyroidism, halts precocious puberty and galactorrhea, and lowers PRL levels. The cause of this syndrome was initially postulated to be increased gonadotropin secretion associated with the massive increase in TSH secretion. However, it appears that TSH, at high concentrations, can activate FSH receptors.

E. McCune-Albright syndrome McCune-Albright syndrome (OMIM#174800. McCune-Albright syndrome; MAS) is classically manifested as a triad of irregular café au lait spots, fibrous dysplasia of long bones with cysts, and precocious puberty. However, hyperthyroidism, adrenal nodules with Cushing syndrome, acromegaly, hyperprolactinemia, hyperparathyroidism, hypophosphatemic hyperphosphaturic rickets, or autonomous endogenous-functioning ovarian cysts (in girls) also occur. Precocious puberty may be central or peripheral; longitudinal studies demonstrate that the history of some patients starts with incomplete precocious puberty and progresses to central precocious puberty. Long-term follow-up of patients with McCune-Albright syndrome reveals a high incidence of pathologic fractures and orthopedic deformities due to the bone cysts, as well as hearing impairment due to the thickening of the temporal area of the skull. Thickening of the skull base can impinge on local brain structures with fatal consequences. Bone growth may be partly mediated by elevated GH secretion and may cease with somatostatin analogue therapy. Patients may have a mutation of Arg_{201} in exon 8 of the gene encoding the Gs alpha subunit that stimulates cAMP formation; this mutation impairs GTPase activity of the alpha subunit and increases adenylate cyclase activity in the affected tissue leading to hormone secretion and the endocrine abnormalities described earlier. This defect originates in somatic rather than germ cells, leading to genetic mosaicism wherever the defective gene product is expressed; thus, a wide variety of tissues can be affected.

Peripheral or Incomplete Contrasexual Precocity in Girls

Excess androgen effect can be caused by premature adrenarche or more significant pathologic conditions such as congenital or non-classic adrenal hyperplasia or adrenal or ovarian tumors. P450c21 adrenal hyperplasia can be diagnosed on the basis of elevated serum 17-hydroxyprogesterone concentrations in the basal or ACTH-stimulated state (other adrenal metabolites may be elevated depending on the defect under investigation) (see Chapter 14). Both adrenal and ovarian tumors are associated with elevation of serum testosterone; adrenal tumors preferentially secrete DHEA. Thus, the source of the tumor may be difficult to differentiate if it produces only testosterone. MRI or CT scanning may be inadequate to diagnose the tumor's organ of origin, and selective venous sampling may be needed. Rare families reported with aromatase deficiency had contrasexual development although not precocious puberty per se. One affected 46,XX individual had virilized genitalia at birth and further virilization at a pubertal age with the additional feature of polycystic ovaries. The patient, in spite of high serum testosterone concentrations, had elevated FSH and LH in the absence of estrogen production.

Variations in Pubertal Development

A. Premature thelarche The term **premature thelarche** denotes unilateral or bilateral breast enlargement without other signs of androgen or estrogen secretion. Patients are usually younger than 3 years of age; the breast enlargement may regress within months or remain until actual pubertal development occurs at a normal age. Areolar development and vaginal mucosal signs of estrogen effect are usually absent or minimal. Premature thelarche may be caused by brief episodes of estrogen secretion from ovarian cysts. Plasma estrogen levels are usually low in this

disorder, perhaps because blood samples are characteristically drawn after the initiating secretory event. However, ultrasensitive estradiol assays show a difference in estrogen production between control girls and those with premature thelarche. Classically, premature thelarche is self-limited and does not lead to central precocious puberty. However, there are reports of progression to central precocious puberty in a minority of cases, and follow-up for such progression is indicated. Other abnormalities in ovulation and menstruation have been reported following premature thelarche. Girls with fluctuating thelarche were reported as having mutations in the *GNAS1* gene, though without other signs of McCune-Albright syndrome.

B. Premature menarche   In rare cases, girls may begin to menstruate at an early age without showing other signs of estrogen effect. An unproved theory suggests that they may be manifesting increased uterine sensitivity to estrogen. In most subjects, menses stop within one to 6 years, and normal pubertal progression occurs thereafter.

C. Premature Adrenarche   The term **premature adrenarche** denotes the early appearance of pubic or axillary hair without other signs of virilization or puberty. This nonprogressive disorder is compatible with the occurrence of other signs of puberty at the normal age for puberty. Premature adrenarche is more common in girls than in boys, and premature adrenarche is usually found in children over 6 years of age. Sometimes premature adrenarche can overlap with the new age limits of puberty in girls. Plasma and urinary DHEAS levels are generally elevated to stage 2 pubertal levels, higher than normally found in this age group but well below those seen with tumors of the adrenal glands. Bone and height ages may be slightly advanced for chronologic age. Patients may have abnormal electroencephalographic tracings without other signs of neurologic dysfunction.

The presenting symptoms of late-onset adrenal hyperplasia may be similar to those of premature adrenarche, and the differential diagnosis may require ACTH stimulation testing (see Chapter 14). This condition may also present in a manner similar to polycystic ovarian syndrome (PCOS) (see Chapter 13). Infants born small for gestational age (SGA) have a predilection to develop premature adrenarche. These girls may also progress to PCOS. Daughters of women with PCOS demonstrate elevated androgen concentrations before the age of puberty, and these daughters may progress themselves to develop PCOS.

D. Adolescent gynecomastia   Up to 75% of boys have transient unilateral or bilateral gynecomastia, usually beginning in stage 2 or 3 of puberty and regressing about 2 years later. Serum estrogen and testosterone concentrations are normal, but the estradiol-testosterone ratio may be elevated and SHBG concentrations may be high. Reassurance is usually all that is required, but some severely affected patients with extremely prominent breast development require reduction mammoplasty if psychologic distress is extreme. After 2 years of gynecomastia, the likelihood of spontaneous regression decreases due to the formation of scar tissue in the breast. Early studies of aromatase inhibitors in the medical treatment of gynecomastia were disappointing.

Some pathologic conditions such as Klinefelter syndrome and the syndromes of incomplete androgen resistance are also associated with gynecomastia; these disorders should be clearly differentiated from the gynecomastia of normal puberty in males. Certain medications list gynecomastia as a side-effect but in most cases it appears to be adipose tissue rather than true glandular breast development. Most of them, including the atypical antipsychotics, raise prolactin levels; hyperprolactinemia does not by itself cause substantial breast development but does cause galactorrhea from activity of the secretory glands. Marijuana use can also increase the risk for gynecomastia.

Differential Diagnosis of Precocious Puberty

The history and physical examination should be directed toward one of the diagnostic possibilities discussed earlier. Serum gonadotropin and sex steroid concentrations are determined in order to distinguish gonadotropin-mediated secondary sexual development (serum gonadotropin and sex steroid levels elevated to pubertal levels) from autonomous endogenous secretion or exogenous administration of gonadal steroids (serum gonadotropin levels suppressed and sex steroid levels elevated). Gonadotropins must be measured in a third-generation assay with pediatric standards, as the customary gonadotropin determinations are too insensitive to reveal the small changes of puberty.

Measurement of testosterone or estradiol must be done in an ultrasensitive assay with pediatric standards or, as with gonadotropin assays, the small changes of pubertal development will be missed. Highly sensitive high performance liquid chromatography tandem mass spectroscopy (HPLC MS/MS) assays are available at national laboratories for this purpose. Third-generation ultrasensitive immunoassays can identify the onset of increased gonadotropin secretion with a single basal unstimulated serum sample. In the past, a GnRH test was required to confirm an increase in LH secretion at puberty because of the overlap of pubertal and prepubertal values of LH in the basal state. These third-generation gonadotropin assays can be applied to urine as well as serum samples; this may someday eliminate the need for GnRH agonist testing or serial serum sampling.

If serum LH concentrations, measured in an assay that cross-reacts with hCG, are quite high in a boy or if a pregnancy screening test (β-hCG) is positive, the likely diagnosis is an extra-pituitary, hCG-secreting tumor. β-LH values are suppressed, so no confusion with hCG levels is possible. If no abdominal or thoracic source of hCG is found, MRI of the brain with particular attention to the hypothalamic-pituitary area is indicated to evaluate the possibility of a germinoma of the pineal gland.

If serum sex steroid levels are very high and gonadotropin levels are low in third-generation assays, an autonomous source of gonadal steroid secretion must be assumed. If plasma gonadotropin and sex steroid levels are in the pubertal range, the most likely diagnosis is complete precocious puberty (Table 15–5).

TABLE 15–5 Differential diagnosis of precocious puberty.

	Serum Gonadotropin Concentrations	LR Response to GnRH or GnRH Agonist	Serum Sex Steroid Concentrations	Gonadal Size	Miscella...
Central (complete or true) precocious puberty					
	Pubertal	Pubertal pattern	Pubertal values	Normal pubertal enlargement of gonads in males	MRI sca... nervo... or oth... lesion...
Peripheral or Incomplete precocious puberty					
Males					
Chorionic gonadotropin secreting tumor	High hCG (or LH in cross-reactive assay), positive pregnancy test, but β-LH (specific assay) is low	High basal hCG or LH that does not rise with GnRH, but β-LH is low and does not rise	High or pubertal values	Slight to moderate enlargement of gonads	Hepatic... MRI s... centr... suspe... hCG, β-hCG... which...
Leydig cell tumor	Prepubertal (low)	Prepubertal or suppressed pattern	Extremely high testosterone	Irregular asymmetric enlargement of testes	
Familial gonadotropin-independent sexual precocity with premature Leydig and germ cell maturation	Prepubertal (low)	Prepubertal or suppressed pattern	Pubertal or higher values	Testes longer than 2.5 cm but smaller than expected for stage of pubertal development	Found i... patte...
Females					
Granulosa cell tumor (follicular cysts may have similar presentation)	Prepubertal (low)	Prepubertal or suppressed pattern	Extremely high estradiol	Ovarian enlargement on physical, CT, MRI, or sonographic examination	Granulo... palpa...
Follicular cyst	Prepubertal (low)	Prepubertal or suppressed pattern of LH secretion. FSH secretion may rise above normal range	High to extremely high estradiol	Cysts may be visible on sonogram	Withdra... estro... recur...

Differentiation between premature thelarche and central precocious puberty is usually accomplished by physical examination, but determination of serum estradiol or gonadotropins may be required. The evaluation of uterine size by ultrasound may also be useful, because premature thelarche causes no increase in uterine volume while central precocious puberty does. Central precocious puberty is also associated with an "endometrial stripe" on ultrasound evaluation. Ovarian size determination is a less useful method of distinguishing between the two possibilities but may reveal pubertal ovarian cysts. As noted earlier, some girls initially thought to have precocious thelarche progress to complete precocious puberty, but there is no way presently to distinguish girls who will progress from those who will not.

The onset of true or complete precocious puberty may indicate the presence of a hypothalamic tumor. Boys more often than girls have central nervous system tumors associated with complete precocious puberty. Skull x-rays are not usually helpful, but CT or MRI scanning is indicated in children with true precocious puberty. The present generation of CT and MRI scanners can make thin cuts through the hypothalamic-pituitary area with good resolution; small hypothalamic hamartomas are now being diagnosed more frequently. Generally, MRI is preferable to CT because of better resolution; the use of contrast may help evaluate possible central nervous system lesions.

Treatment of Precocious Puberty

A. Central precocious puberty   In the past, medical treatment of true precocious puberty involved medroxyprogesterone acetate or cyproterone acetate—progestational agents that reduce gonadotropin secretion by negative feedback.

However, current treatment for central precocious puberty, either idiopathic or due to a central nervous system lesion, involves the use of GnRH agonists that suppress sexual maturation and decrease growth rate and skeletal maturation. Chronic administration of highly potent and long-acting analogs of GnRH has been shown to down regulate GnRH receptors and reduce pituitary gland response to GnRH, thereby causing decreased secretion of gonadotropin and sex steroids and rapidly stopping the progression of sexual precocity. In girls, pubertal enlargement of the ovaries and uterus reverts toward the prepubertal state, and the multicystic appearance of the pubertal ovary regresses as well. This suppressive effect is reversed after therapy is discontinued and puberty progresses. GnRH agonist treatment has been successful for idiopathic precocious puberty and precocious puberty caused by hamartomas of the tuber cinereum, neoplasms of the central nervous system, or long-term androgen exposure. The GnRH agonists were originally given by daily subcutaneous injection or intranasal insufflation. Injection of GnRH agonists in depot preparations every 4 weeks, or every 3 months in a more recent preparation, or in an implantable silastic pellet which is left in place for one or 2 years has now made treatment much more effective. The FDA has approved histrelin, a GnRH agonist administered on a daily basis, and leuprolide acetate, a long-acting GnRH agonist, administered every 28 days, or every 3 months, for the management of precocious puberty. Individual monitoring

is essential to ascertain gonadotropin suppression. An implantable depot preparation of the GnRH agonist histrelin may be left in for 1 to 2 years without the requirement for frequent monitoring. Complete suppression of pubertal devlopment is necessary because an incompletely suppressed patient may appear to have arrested pubertal development while actually secreting low but significant levels of sex steroids. Without clinical suppression, bone age advances while the growth rate is decreased, leading to even shorter adult stature. However gonadotropin secretion need not be completely suppressed for effective management of precocious puberty as there are reports in which patients were clinically controlled while gonadotropins values were not suppressed to the prepubertal range. Side-effects of these agonists have generally been limited to allergic skin reactions and elevation of immunoglobulins directed against GnRH. Significant anaphylactic reactions to an injection have rarely been reported. Decrease in bone mineral density is a potential side-effect of GnRH agonists; increased dietary calcium and supplementation of vitamin D intake is wise especially because there are frequent deficiencies in the majority of children.

Growth velocity decreases within 5 months after the start of therapy, and rapid bone age advancement decreases to a rate below the increase in chronologic age. Without therapy, adult height in patients with central precocious puberty approaches 152 cm in girls and 155 to 164 cm in boys. The first patients treated with GnRH agonists have now reached adult height: the girls have a mean height of 157 cm and the boys a mean height of 164 cm—a definite improvement over the untreated state. The shortest height prognosis is found in children with an early onset of precocious puberty who are not treated. The best outcomes of therapy are noted when diagnosis and therapy are achieved early, and as earlier diagnosis is now made in children with central precocious puberty and earlier therapy is offered, better results are expected in the future. Menarche and even pregnancy are reported after the discontinuation of therapy in girls, indicating a reversion to normal pubertal endocrine function following GnRH agonist treatment. The new lower age limits of normal pubertal development in girls have led clinicians to reassess the criteria used to identify appropriate candidates for therapy. Patients without significant elevation of serum estrogen, who have a predicted height appropriate for family and who have slowly progressing variants without early menarche, will likely achieve an appropriate adult height without therapy.

The somatic changes or menses frighten some children and may make them the object of ridicule. These patients do not experience social maturation to match their physical development, although their peers, teachers, and relatives tend to treat them as if they were older because of their large size. Thus, supportive counseling must be offered to both patient and family. Children with precocious puberty are more often sexually targeted and subsequently abused, so appropriate precautions are necessary.

B. Peripheral or Incomplete precocious puberty   Treatment of the disorders discussed earlier under peripheral precocious puberty is directed toward the underlying tumor or abnormality rather than toward the signs of precocious puberty. If the primary

cause is controlled, signs of sexual development are halted and may even regress.

Males with familial Leydig and germ cell maturation do not initially respond to GnRH agonist therapy, but some have improved with medroxyprogesterone acetate. Affected boys were successfully treated with ketoconazole, an antifungal agent that can block 17,20-lyase activity and, therefore, decrease testosterone production. However, ketoconazole also suppresses adrenal function as a significant side-effect. After initial control with ketoconazole, the boys developed central precocious puberty, because prolonged exposure to androgens led to maturation of their hypothalamic-pituitary axis; treatment with a GnRH agonist then effectively halted this pubertal progression.

Newer therapy for McCune-Albright syndrome in girls is a combination of testolactone (an aromatase inhibitor) and spironolactone (which acts as an antiandrogen). Long-term follow-up demonstrated some decrease in menses, improvement in growth patterns and bone age advancement. Some escaped from control, necessitating the addition of a GnRH agonist, which then suppressed pubertal development. The combination of tamoxifen or fulvesent, estrogen receptor blockers, with aromatase inhibitors (such as letrazole) has demonstrated some success in decreasing uterine bleeding and advancement in bone age in clinical trials. Girls with recurrent estrogen-secreting ovarian cyst formation may have a decreased incidence of cysts with medroxyprogesterone acetate therapy. A GnRH agonist may also be effective in such cases. Surgical removal of ovarian cysts may be unnecessary if such medical therapy is utilized first.

Precocious thelarche or adrenarche requires no treatment, as both are self-limited benign conditions. No therapy has been reported for premature menarche. Severe, persistent cases of adolescent gynecomastia have reportedly been treated successfully by testolactone and dihydrotestosterone heptanoate, although surgical removal of breast tissue is often necessary, as noted earlier.

REFERENCES

General

Albrecht L, Styne D. Laboratory testing of gonadal steroids in children. *Pediatr Endocrinol Rev*. 2007;5(Suppl 1):599. [PMID: 18167469]

Cousminer DL, Berry DJ, Timpson NJ, et al. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Hum Mol Genet*. 2013;22:2735. [PMID: 23449627]

Finlayson CA, Styne DM, Jameson JL. Endocrinology of Sexual Maturation and Puberty. In: Jameson JL, De Groot LJ, eds. *Endocrinology: Adult and Pediatric*. 7th ed. Philadelphia, PA: Elsevier; 2015:2119-2129.

Gore AC, Chappell VA, Fenton SE, et al. Executive summary to EDC-2: The Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocr Rev*. 2015;36:593. [PMID: 26544531]

Greulich WW, Pyle SI. *Radiographic Atlas of Skeletal Development of the Hand and Wrist*. 2nd ed. Stanford. CA: Stanford University Press; 1959.

Herman-Giddens ME, Slora EJ, Wasserman RC, et al. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics*. 1997;99:505. [PMID: 9093289]

Himes JH, Park K, Styne D. Menarche and assessment of body mass index in adolescent girls. *J Pediatr*. 2009;155:393. [PMID: 199559445]

Kaiser UB, Kuohung W. KiSS-1 and GPR54 as new players in gonadotropin regulation and puberty. *Endocrine*. 2005;26:277. [PMID: 12944565]

Lee JM, Wasserman R, Kaciroti N, et al. Timing of puberty in overweight versus obese boys. *Pediatrics*. 2016;137:1. [PMID: 26817933]

Lee Y, Styne D. Influences on the onset and tempo of puberty in human beings and implications for adolescent psychological development. *Horm Behav*. 2013;64:250. [PMID: 2398669]

Main KM, Skakkebaek NE, Toppari J. Cryptorchidism as part of the testicular dysgenesis syndrome: the environmental connection. *Endocr Dev*. 2009;14:167. [PMID: 119293583]

Palmert MR, Dunkel L, Witchel SF. Puberty and its disorders in the male. In: Sperling MA, ed. *Pediatric Endocrinology*. 4th ed. Philadelphia, PA: Elsevier; 2014:697–733.

Rosenfield RL, Cooke DW, Radovick S. Puberty and its disorders in the female. In: Sperling MA, eds. *Pediatric Endocrinology*. 4th ed. Philadelphia, PA: Elsevier; 2014:569–663.

Rosenfeld RL, Lipton RB, Drum ML. Thelarche, pubarche, and menarche attainment in children with normal and elevated body mass index. *Pediatrics*. 2009;123:84. [PMID: 19117864]

Rosenthal SM. Approach to the patient: transgender youth: endocrine considerations. *J Clin Endocrinol Metab*. 2014;99:4379. [PMID: 25140398]

Styne DM. Puberty, obesity and ethnicity. *Trends Endocrinol Metab*. 2004;15:472. [PMID: 15541646]

Styne DM, Grumbach MM. Physiology and disorders of puberty. In: Melmed S, Polonsky KS, Larsen PR, Kronenberg HM, eds. *Williams Textbook of Endocrinology*. 13th ed. Philadelphia, PA: Saunders Elsevier; 2015:1074–1218.

Physical Changes Associated with Puberty

Bayley N, Pinneau SF. Tables for predicting adult height from skeletal age: revised for use with the Greulich-Pyle standards. *J Pediatr*. 1952;40:423. [PMID: 14918032]

Loomba-Albrecht LA, Styne DM. Effect of puberty on body composition. *Curr Opin Endocrinol Diabetes Obes*. 2009;16:10. [PMID: 19115520]

Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in girls. *Arch Dis Child*. 1969;44:291. [PMID: 5785179]

Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*. 1970;45:13. [PMID: 5440182]

Delayed Puberty and Sexual Infantilism

Achermann JC, Hughes IA. Disorders of sex development. In: Melmed S, Polonsky KS, Larsen PR, Kronenberg HM, eds. *Williams Textbook of Endocrinology*. 12th ed. Philadelphia, PA: Saunders, Elsevier; 2015:868-934.

Bondy CA, Turner Syndrome Study Group. Care of girls and women with Turner syndrome: a guideline of the Turner Syndrome Study Group. *J Clin Endocrinol Metab*. 2007;92:10. [PMID: 17047017]

Davenport ML. Approach to the patient with Turner syndrome. *J Clin Endocrinol Metab*. 2010;95:1487. [PMID: 20375216]

Harrington J, Palmert MR. Clinical review: distinguishing constitutional delay of growth and puberty from isolated hypogonadotropic hypogonadism: critical appraisal of available diagnostic tests. *J Clin Endocrinol Metab*. 2012;97:3056. [PMID: 22723321]

Johansson T, Ritzen EM. Very long-term follow-up of girls with early and late menarche. *Endocr Dev*. 2005;8:126. [PMID: 15722621]

Jona K, Coutifaris C, Gracia CR, et al. Future directions in oncofertility and fertility preservation: a report from the 2011 Oncofertility Consortium Conference. *J Adolesc Young Adult Oncol*. 2013;2:25. [PMID: 23610740]

Mountjoy M, Sundgot-Borgen J, Burke L, et al. The IOC consensus statement: beyond the Female Athlete Triad–Relative Energy Deficiency in Sport (RED-S). *Br J Sports Med*. 2014;48:491. [PMID: 24620037]

Nabhan ZM, Dimeglio LA, Qi R, Perkins SM, Eugster EA. Conjugated oral versus transdermal estrogen replacement in girls with Turner syndrome: a pilot comparative study. *J Clin Endocrinol Metab*. 2009;94:2009. [PMID: 19318455]

Palmert MR, Dunkel L. Clinical practice. Delayed puberty. *N Engl J Med*. 2012;366:443. [PMID: 22296078]

Rosenfield RL. Clinical review: adolescent anovulation: maturational mechanisms and implications. *J Clin Endocrinol Metab*. 2013;98:3572. [PMID: 23913942]

Sedlmeyer IL, Palmert MR. Delayed puberty: analysis of a large case series from an academic center. *J Clin Endocrinol Metab*. 2002;87:1613. [PMID: 11932291]

Sexual Precocity

Abreu AP, Dauber A, Macedo DB, et al. Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N Engl J Med*. 2013;368:2467. [PMID: 23738509]

Bar A. Bayley-Pinneau method of height prediction in girls with central precocious puberty: correlation with adult height. *J Pediatr*. 1995;126:955. [PMID: 7776106]

Carel JC, Eugster EA, Rogol A. Consensus statement on the use of gonadotropin-releasing hormone analogs in children. *Pediatrics*. 2009;123:e752. [PMID: 19332438]

Henrichs KL, McCauley HL, Miller E, et al. Early menarche and childhood adversities in a nationally representative sample. *Int J Pediatr Endocrinol*. 2014;2014:14. [PMID: 25089128]

Hughes IA. Releasing the brake on puberty. *N Engl J Med*. 2013;368:2513. [PMID: 23738512]

Latronico AC, Brito VN, Carel JC. Causes, diagnosis, and treatment of central precocious puberty. *Lancet Diabetes Endocrinol*. 2016;4:265. [PMID: 26852255]

Klein KO, Barnes KM, Jones JV, et al. Increased final height in precocious puberty after long-term treatment with LHRH agonists: the National Institutes of Health experience. *J Clin Endocrinol Metab*. 2001;86:4711. [PMID: 11600530]

Legro RS, Arslanian SA, Ehrmann DA, et al. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2013;98:4565. [PMID: 24151290]

Lenz AM, Shulman D, Eugster EA, et al. Bicalutamide and third generation aromatase inhibitors in testotoxicosis. *Pediatrics*. 2010;126:e728. [PMID: 20713483]

Mul D, Bertelloni S, Carel JC, et al. Effect of gonadotropin-releasing hormone agonist treatment in boys with central precocious puberty: final height results. *Horm Res*. 2002;58:1. [PMID: 12169774]

Palmert MR, Malin HV, Boepple PA. Unsustained or slowly progressive puberty in young girls: initial presentation and long-term follow-up of 20 untreated patients. *J Clin Endocrinol Metab*. 1999;84:415. [PMID: 10022394]

Reiter EO, Lindberg A, Ranke MB, et al. The KIGS experience with the addition of gonadotropin-releasing hormone agonists to growth hormone (GH) treatment of children with idiopathic GH deficiency. *Horm Res*. 2003;60 (suppl 1):68. [PMID: 12955021]

Tanaka T, Niimi H, Matsuo N, et al. Results of long-term follow-up after treatment of central precocious puberty with leuprorelin acetate: evaluation of effectiveness of treatment and recovery of gonadal function. The TAP-144-SR Japanese Study Group on Central Precocious Puberty. *J Clin Endocrinol Metab*. 2005;90:1371. [PMID: 15598675]

Teilmann G, Petersen JH, Gormsen M, et al. Early puberty in internationally adopted girls: hormonal and clinical markers of puberty in 276 girls examined biannually over two years. *Horm Res*. 2009;72:236. [PMID: 19786795]

Voutilainen R, Jaaskelainen J. Premature adrenarche: etiology, clinical findings, and consequences. *J Steroid Biochem Mol Biol*. 2015;145:226. [PMID: 24923732]

Weise M, Flor A, Barnes KM, Cutler GB Jr, Baron J. Determinants of growth during gonadotropin-releasing hormone analog therapy for precocious puberty. *J Clin Endocrinol Metab*. 2004;89:103. [PMID: 14715835]

The Endocrinology of Pregnancy

Bansari Patel, MD, Joshua F. Nitsche, MD, PhD, and Robert N. Taylor, MD, PhD

ACTH	Adrenocorticotropin hormone	**hPL**	Human placental lactogen
AFP	Alpha-fetoprotein	**IGF**	Insulin-like growth factor
AMH	Anti-Müllerian hormone	**LH**	Luteinizing hormone
Ang 1-7	Angiotensin 1-7	**NSAID**	Nonsteroidal anti-inflammatory drug
BMI	Body mass index	**OTR**	Oxytocin receptor
BV	Bacterial vaginosis	**PDGF**	Platelet-derived growth factor
CAP	Contraction-associated protein	**PG**	Prostaglandin
CRH	Corticotropin-releasing hormone	**PGR-A and -B**	Progesterone receptor A and B
Cx43	Connexin 43		
DHEA	Dehydroepiandrosterone	**PIGF**	Placental growth factor
EGF	Epidermal growth factor	**PRL**	Prolactin
fFN	Fetal fibronectin	**PROM**	Premature rupture of membranes
FGF	Fibroblast growth factor	**PTB**	Preterm birth
GDM	Gestational diabetes mellitus	**sEng**	Soluble endoglin
GH	Growth hormone	**sFlt-1**	Soluble fms-like tyrosine kinase 1
GnRH	Gonadotropin-releasing hormone	**SHBG**	Sex hormone–binding globulin
GTN	Gestational trophoblastic neoplasia	**TRH**	Thyrotropin-releasing hormone
hCG	human chorionic gonadotropin	**TSH**	Thyroid-stimulating hormone
HELLP syndrome	Hemolysis, elevated liver enzymes, and low platelets	**VEGF**	Vascular endothelial growth factor

Throughout pregnancy, the fetal-placental unit secretes protein and steroid hormones and eicosanoids that alter the function of every endocrine gland in the mother's body. Both clinically and in the laboratory, pregnancy can mimic hyperthyroidism, Cushing disease, pituitary adenoma, diabetes mellitus, polycystic ovary syndrome, and more.

The endocrine changes associated with pregnancy are adaptive, allowing the mother to nurture the developing fetus. Although maternal reserves are usually adequate, in cases of gestational diabetes or hypertensive disease of pregnancy, a woman may develop overt signs of disease as a direct result of pregnancy.

Aside from creating a satisfactory nutritive environment for fetal development, the placenta serves as an endocrine, respiratory, alimentary, and excretory organ. Measurements of fetal-placental products in the maternal serum provide one means of assessing fetal well-being. This chapter will consider the changes in maternal endocrine function in pregnancy and during-parturition as well as fetal endocrine development. The chapter concludes with a discussion of some endocrine disorders complicating pregnancy.

CONCEPTION AND IMPLANTATION

Fertilization

In fertile women, ovulation occurs approximately 12 to 16 days after the onset of the previous menses. The ovum must be

fertilized within 24 to 48 hours if conception is to result. For about 48 hours around ovulation, cervical mucus is copious, non-viscous, slightly alkaline, and forms a gel matrix that acts as a filter and conduit for sperm. Sperm begin appearing in the outer third of the fallopian tube (the ampulla) 5 to 10 minutes after coitus and continue to migrate to this location from the cervix for about 24 to 48 hours. Of the 200×10^6 sperm that are deposited in the vaginal fornices, only approximately 200 reach the distal tube. Fertilization normally occurs in the ampulla.

Implantation and hCG Production

After fertilization and zygote transport into the endometrial cavity, blastocyst invasion of the uterus occurs during a specific window of implantation 8 to 10 days after ovulation. Vitronectin, an alpha-v-beta-3 integrin receptor ligand, serves as one of several links between the maternal and embryonic epithelia. Two layers of placental epithelial cells, cytotrophoblasts and syncytiotrophoblasts, develop after the blastocyst invades the endometrium (Figure 16–1).
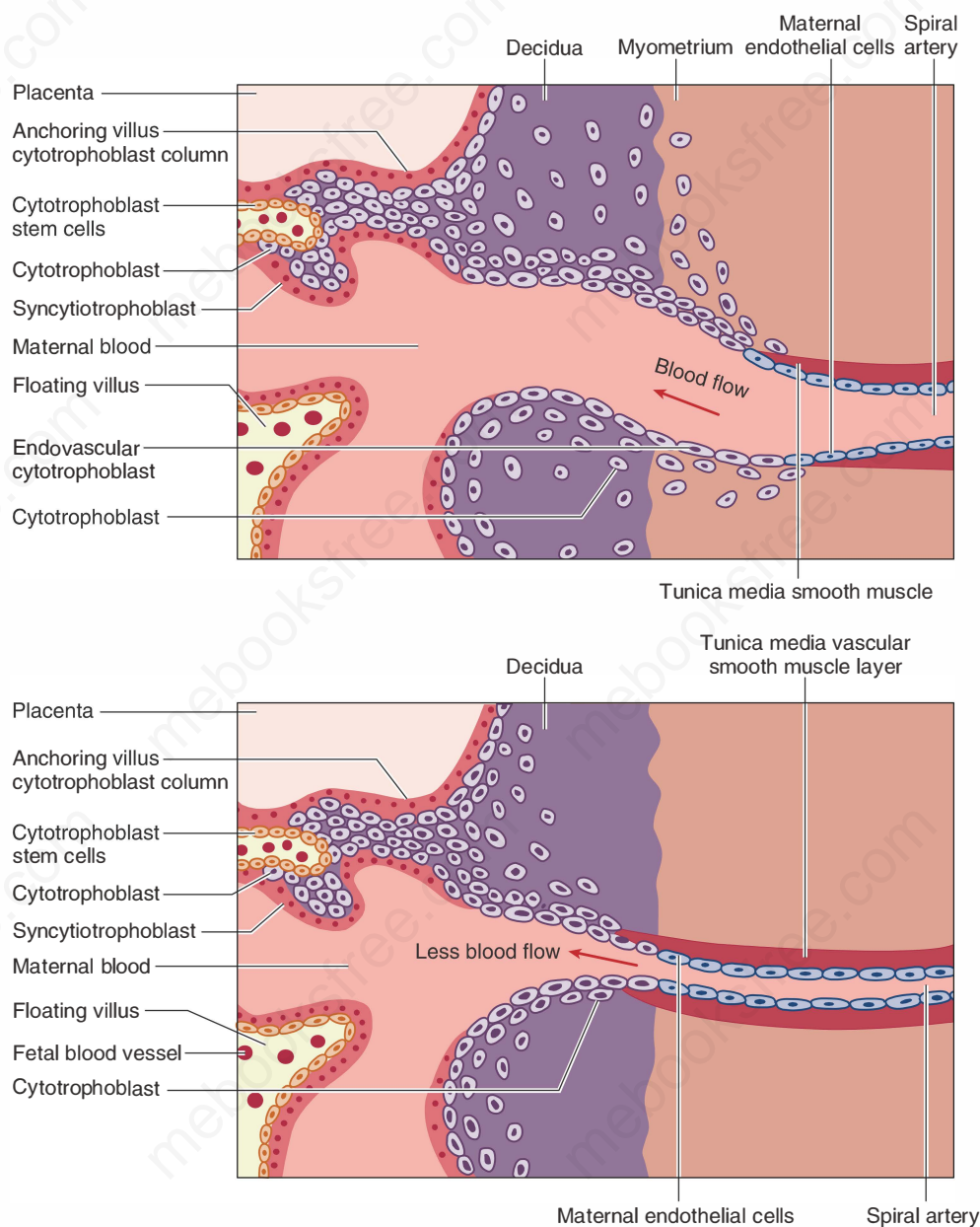
FIGURE 16–1 Microanatomy of the human placental bed. Oxygen, nutrients, and waste products are exchanged between the fetus and the mother across this interface. In normal placental development (upper panel), invasive cytotrophoblasts of fetal origin transform the maternal spiral arteries, displacing maternal endothelial cells and remodeling the vessels from small-caliber resistance vessels to high-caliber capacitance vessels capable of placental perfusion adequate to sustain fetal growth. In preeclampsia and some cases of fetal growth restriction, cytotrophoblasts fail to adopt an invasive endothelial phenotype. Instead, invasion of the spiral arteries is shallow and they remain small-caliber resistance vessels (lower panel). This can lead to placental ischemia. (Reproduced with permission from Karumanchi SA, Maynard SE, Stillman IE, et al. Preeclampsia: a renal perspective. *Kidney Int*. 2005 Jun;67(6):2101-2113.)

Columns of invading cytotrophoblasts anchor the placenta to the endometrium. The differentiated syncytiotrophoblast, derived from fusion of cytotrophoblasts, is in direct contact with the maternal circulation. The syncytiotrophoblast is the major source of hormone production, containing the cellular machinery needed for synthesis and secretion of both steroid and polypeptide hormones.

In most spontaneously conceived pregnancies, the dates of ovulation and implantation are not known. Weeks of gestation (*gestational age*) are by convention calculated from the first day of the last menstrual period. Within 24 hours after implantation, or at about 3 weeks of gestation, human chorionic gonadotropin (hCG), produced by syncytiotrophoblasts (see Figure 16–1), is detectable in maternal serum. Under the influence of increasing hCG production, the corpus luteum secretes progesterone, estradiol, and relaxin in increasing quantities.

Ovarian Hormones of Pregnancy

The hormones produced by the corpus luteum include progesterone, 17-hydroxyprogesterone, relaxin, and estradiol. The indispensability of the corpus luteum in early pregnancy has been demonstrated by ablation studies, in which luteectomy or oophorectomy before 42 days of gestation results in precipitous decreases in levels of serum progesterone and estradiol, followed by abortion. Exogenous progesterone will prevent abortion, proving that progesterone alone is required for maintenance of early pregnancy. After about the seventh gestational week, the corpus luteum can be removed without subsequent abortion owing to compensatory progesterone production by the placenta.

Because the placenta does not express the 17α-hydroxylase enzyme (P450C17), it cannot produce appreciable amounts of 17-hydroxyprogesterone; thus, this steroid provides a marker of corpus luteum function. As shown in Figure 16–2, the serum concentrations of estrogens and total progesterone exhibit a steady increase, but the concentration of 17-hydroxyprogesterone rises and then declines to low levels that persist for the duration of the pregnancy. Another marker of corpus luteum function is the polypeptide hormone relaxin, a protein with a molecular mass of about 6000. Pharmacologically, relaxin ripens the cervix, softens the pubic symphysis, promotes decidual angiogenesis, and acts synergistically with progesterone to inhibit uterine contractions.

Symptoms and Signs of Pregnancy

Breast tenderness, fatigue, nausea, absence of menstruation, softening of the uterus, and a sustained elevation of basal body temperature are mostly attributable to hormone production by the corpus luteum and developing placenta.

FETAL-PLACENTAL-DECIDUAL UNIT

The function of the placenta is to establish effective communication between the mother and the developing fetus while maintaining the immune and genetic integrity of both individuals. Initially, the placenta functions autonomously. By the end of the first trimester, however, the fetal endocrine system is sufficiently developed to complement placental function and to provide some hormone precursors to the placenta. From this time, it is useful to consider the endocrine conceptus as the fetal-placental unit.

The decidua is the endometrium of pregnancy. Decidual cells are capable of synthesizing a variety of polypeptide hormones, including relaxin, prolactin (PRL), and a variety of paracrine factors, in particular vascular endothelial growth factor (VEGF) and insulin-like growth factor (IGF)-binding protein 1. The precise function of the decidua as an endocrine organ has not been established, but its role as a source of prostaglandins during labor is certain (see "Endocrine Control of Parturition," later).

POLYPEPTIDE HORMONES

Human Chorionic Gonadotropin

The first marker of trophoblast differentiation and the first measurable product of the placenta is hCG, a glycoprotein consisting of 237 amino acids. It is similar in structure to the pituitary glycoprotein hormones in that it consists of two chains: a common alpha chain and a specific beta chain, which determines receptor interaction and ultimate biologic specificity. The alpha chain is identical in sequence to the alpha chains of thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), and luteinizing hormone (LH). The beta chain has 67% sequence homology with LH and an additional 30 carboxy terminal amino acids not found in LH beta.

In early pregnancy (up to 6 weeks), the concentration of hCG doubles every 1.7 to 2 days, and serial measurements provide a sensitive index of early trophoblast function and viability. Maternal plasma hCG peaks during the tenth gestational week and then declines gradually in the third trimester. Peak concentrations correlate temporally with the establishment of maternal blood flow in the intervillous space (see Figure 16–1, upper panel).

The long plasma half-life of hCG (24 hours) allows the tiny mass of cells comprising the blastocyst to produce sufficient hormone to be detected in the peripheral circulation within 24 hours of implantation. Thus, pregnancy can be diagnosed several days before symptoms occur or a menstrual period has been missed. Antibodies to the unique beta-carboxyl terminal segment of hCG do not cross-react significantly with any of the pituitary glycoproteins.

hCG is also produced in gestational trophoblastic neoplasia (GTN) by hydatidiform mole and choriocarcinoma, and the concentration of hCG-beta is used as a tumor marker for diagnosis and for monitoring the success of chemotherapy in these disorders. Women with very high hCG levels due to GTN may become clinically hyperthyroid due to the action of hCG on TSH receptors; they can revert to euthyroidism as hCG is reduced during chemotherapy.

Human Placental Lactogen

A second placental polypeptide hormone is human placental lactogen (hPL). hPL is produced by early trophoblasts, but detectable

SYSTEM	HORMONE	PATTERN	AVERAGE PEAK CONCENTRATION (TIME)
Placenta and corpus luteum	Progesterone	Rises to term	190 ng/mL (552 nmol/L) (term)
	17-Hydroxy-progester-one	Peaks at 5 wk, then declines	6 ng/mL (19 nmol/L) (5 wk)
Adrenal	Cortisol	Increases to 3 times prepreg-nancy values at term	300 ng/mL (0.83 µmol/L) (term)
	Aldosterone	Plateaus at 34 wk with small rise near term	100 ng/mL (277 nmol/L) (term)
	DOC	Increases to 10 times prepreg-nancy value at term	1200 pg/mL (3.48 nmol/L) (term)
Thyroid	Total T$_4$	Increases during first trimester, then plateaus	150 ng/mL (193 pmol/L)
	Free T$_4$	Unchanged	30 pg/mL (38.6 pmol/L)
	Total T$_3$	Increases during first trimester, then plateaus	2 ng/mL (3.1 nmol/L)
	Free T$_3$	Unchanged	4 pg/mL (5.1 pmol/L)
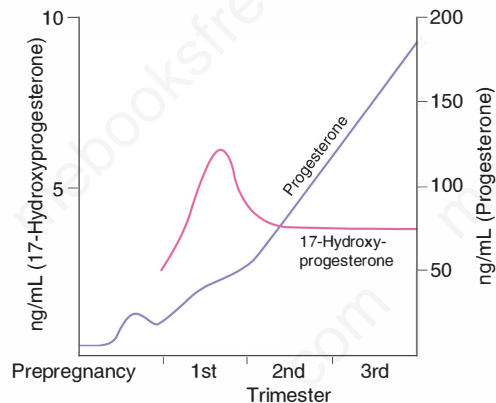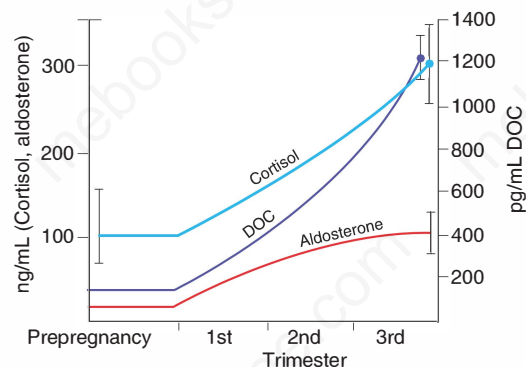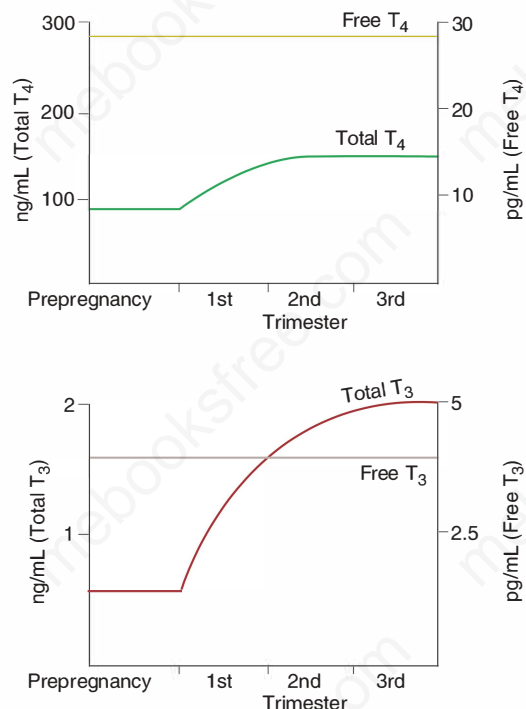
FIGURE 16–2 Maternal serum hormone changes during pregnancy.

SYSTEM	HORMONE	PATTERN	AVERAGE PEAK CONCENTRATION (TIME)
Anterior pituitary	GH	Unchanged	
	LH, FSH	Low, basal levels	
	ACTH	Unchanged	
	TSH	Reaches nadir during first trimester, then plateaus	
	PRL	Rise to term	200 ng/mL (9 nmol/L) (term)

Placental proteins	hCG	Peaks at 10 wk, then decreases to a lower plateau	5 µg/mL (0.2 µmol/L) (end of first trimester)
	hPL	Rises with placental weight	5-25 µg/mL (term) (0.2-1.0 µmol/L)
	CRH	Rises acutely about 20 d before delivery	1-3 ng/mL (2-6 nmol/L) at term

Fetoplacental estrogens	Estriol	Increases to term	15-17 ng/mL (55-62 nmol/L) (term)
	Estradiol	Increases to term	12-15 ng/mL (42-52 nmol/L) (term)
	Estrone	Increases to term	5-7 ng/mL (18.5-26 nmol/L) (term)

Fetoplacental androgens	Testosterone	Rises to 10 times pre-pregnancy values	2000 pg/mL (6.9 nmol/L) (term)
	DHEA	Falls during pregnancy	5 ng/mL (17.3 nmol/L) (pre-pregnancy)
	Androstenedione	Small increase	2.6 ng/mL (9.0 nmol/L) (term)
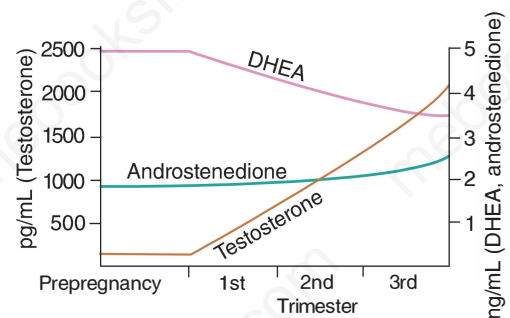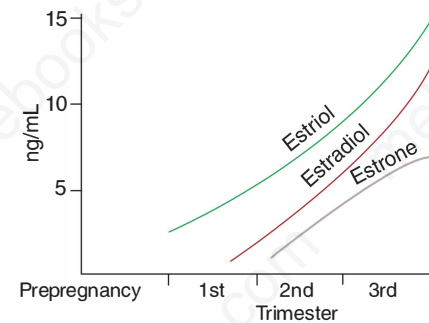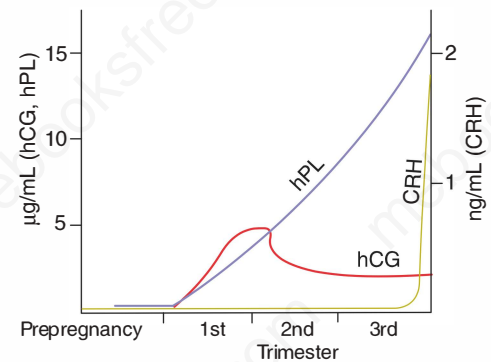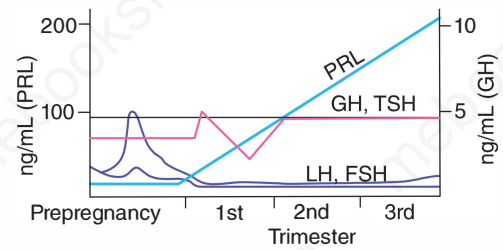
FIGURE 16–2 (*Continued*)

serum concentrations are not reached until 4 to 5 gestational weeks (see Figure 16–2). hPL is a protein of 190 amino acids whose primary, secondary, and tertiary structures are similar to those of growth hormone (GH) and PRL. hPL is diabetogenic and lactogenic, but it has minimal growth-promoting activity as measured by standard GH bioassays.

The physiologic role of hPL during pregnancy remains controversial, and normal pregnancy without detectable hPL production has been reported. Although not clearly shown to be a mammotropic agent, hPL contributes to altered maternal glucose metabolism and mobilization of free fatty acids; causes a hyperinsulinemic response to glucose loads; appears to directly stimulate pancreatic islet insulin secretion; and contributes to the peripheral insulin resistance characteristic of pregnancy. Along with prolonged fasting and insulin-induced hypoglycemia, pre-beta-HDL and apoprotein A-I are two factors that stimulate release of hPL. hPL production is roughly proportionate to placental mass. Actual production rates may reach as much as 1 to 1.5 g/d at term.

Serum hPL concentrations were used historically as a clinical indicator of the health of the placenta, but the range of normal values was wide, and serial determinations were necessary; they are therefore no longer used clinically.

Other Chorionic Peptide Hormones and Growth Factors

Other chorionic peptides have been identified, but their functions remain poorly defined. Activin, inhibin, corticotropin-releasing hormone, and multiple peptide growth factors, including fibroblast growth factor (FGF), epidermal growth factor (EGF), platelet-derived growth factor (PDGF), and the IGFs all have been isolated from placental tissue. Placental growth factor (PlGF) and the related VEGF are suggested to play a role in placental angiogenesis, preeclampsia, and fetal growth.

STEROID HORMONES

In contrast to the impressive synthetic capability exhibited in the production of placental proteins, the placenta does not have the ability to synthesize steroids de novo. However, the trophoblasts have remarkable capacity to efficiently interconvert steroids derived from maternal or fetal precursors. This activity is demonstrable even in the early blastocyst, and by the seventh gestational week, when the corpus luteum undergoes involution, the placenta becomes the dominant source of steroid hormones.

Progesterone

The placenta relies on maternal cholesterol as its substrate for progesterone production. Fetal death has no immediate influence on progesterone production, suggesting that the fetus is a negligible source of substrate. Enzymes in the placenta cleave the cholesterol side chain, yielding pregnenolone, which in turn is isomerized to progesterone; 250 to 350 mg of progesterone are produced daily by the third-trimester placenta, and most enters the maternal circulation (see Figure 16–2). Whereas exogenous

hCG stimulates progesterone production in early pregnancy, hypophysectomy, adrenalectomy, or oophorectomy have no effect on progesterone levels after the luteo-placental shift at 7 to 9 weeks of gestation. Likewise, the administration of adrenocorticotropin (ACTH) or cortisol does not influence placental progesterone secretion.

Progesterone is necessary for establishment and maintenance of pregnancy. Insufficient production of progesterone may contribute to failure of implantation, recurrent pregnancy loss, and preterm delivery. Progesterone, along with relaxin and nitric oxide, maintains uterine quiescence during pregnancy. Clinical trials have indicated that the administration of 17α-hydroxyprogesterone caproate can reliably prevent preterm delivery in high-risk pregnancies. Progesterone also inhibits T cell–mediated allograft rejection. Thus, high local concentrations of progesterone may contribute to immunologic tolerance by the uterus of invading trophoblast tissue from the semiallogeneic fetus.

Estrogens

Estrogen production by the placenta also depends on circulating precursors, but in this case both fetal and maternal steroids are important sources. Most of the placental estrogens are derived from fetal androgens, primarily dehydroepiandrosterone (DHEA) sulfate. Fetal DHEA sulfate, produced mainly by the fetal adrenal, is converted by placental sulfatase to free DHEA and then, through enzymatic pathways common to steroid-producing tissues, to androstenedione and testosterone. These androgens are finally aromatized by the placenta to estrone and estradiol, respectively. 17α-Hydroxysteroid dehydrogenase type II prevents fetal exposure to potent estrogens by catalyzing the conversion of estradiol to less potent estrone.

Most fetal DHEA sulfate is metabolized to produce a third estrogen: estriol. Estriol is a weak estrogen with 1/10 the potency of estrone and 1/100 the potency of estradiol. Serum estrone and estradiol concentrations are increased during pregnancy about 50-fold over their maximal prepregnancy values, but estriol increases approximately 1000-fold. The substrate for the reaction is 16α-hydroxy-DHEA sulfate produced in the fetal adrenal and liver, not in maternal or placental tissues. The final steps of desulfation and aromatization to estriol occur in the placenta. Maternal serum or urinary estriol measurements, unlike measurements of progesterone or hPL, reflect fetal as well as placental function. Rising serum or urinary estriol concentrations were once used as biochemical indicators of fetal well-being (see Figure 16–2). Decreased estriol production can result from congenital derangements or iatrogenic intervention. Maternal estriol remains low in pregnancies with placental sulfatase deficiency and in cases of fetal anencephaly. In the first case, DHEA sulfate cannot be hydrolyzed; in the second, little fetal DHEA is produced because fetal adrenal growth stimulation by ACTH is lacking. Maternal administration of glucocorticoids also inhibits fetal ACTH and lowers maternal estriol. Administration of DHEA to the mother during a healthy pregnancy increases estriol production. Placental corticotropin-releasing hormone (CRH) also may be an important regulator of fetal adrenal DHEA sulfate secretion.

Modern methods of screening for fetal chromosomal aneuploidy, particularly trisomy 21 (Down syndrome), utilize circulating biochemical markers. Screening by maternal age alone (>35 years) led to the prenatal identification of only about 25% of aneuploid fetuses. As an aneuploid chromosome complement affects both fetal and placental tissues, their protein and steroid profiles can be informative. A combination of alpha-fetoprotein (AFP), hCG, inhibin-A and unconjugated estriol concentrations, measured in maternal serum between 15 and 18 weeks of gestation, can be used to identify fetal Down syndrome and trisomy 18 with a detection rate of >80% over all age groups.

MATERNAL ADAPTATION TO PREGNANCY

As a successful *parasite*, the fetal-placental unit manipulates the maternal *host* for its own gain but normally avoids imposing excessive stress that would jeopardize the pregnancy. The prodigious production of polypeptide and steroid hormones by the fetal-placental unit directly or indirectly results in physiologic adaptations of virtually every maternal organ system. These alterations are summarized in Figure 16–3. Most of the commonly measured maternal endocrine function tests are radically changed. In some cases, true physiologic alteration has occurred; in others, the changes are due to increased production of hormone–binding proteins by the liver or to decreased serum levels of albumin due to the mild dilutional anemia of pregnancy. In addition, some endocrine changes are mediated by altered clearance rates due to increased glomerular filtration, decreased hepatic excretion, or metabolic clearance of steroid and protein hormones by the placenta. The changes in endocrine function tests are summarized in Table 16–1. Failure to recognize normal pregnancy-induced alterations in endocrine function tests can lead to unnecessary diagnostic tests and interventions that may be seriously detrimental to mother and fetus.

Maternal Pituitary Gland

The mother's anterior pituitary gland hormones have little influence on pregnancy after implantation has occurred. The gland itself enlarges by about one-third, with the major component of this increase being hyperplasia of the lactotrophs in response to the high plasma estrogens. PRL, the product of the lactotrophs, is the only anterior pituitary hormone that rises progressively during pregnancy and peaks at the time of delivery. In spite of its high serum concentrations, pulsatile release of PRL and food-induced and nocturnal increases persist. Hence, the normal neuroendocrine regulatory mechanisms appear to be intact in the maternal adenohypophysis. In nonlactating women, maternal PRL decreases to pregestational levels within 3 months of delivery. Pituitary ACTH and TSH secretion remain unchanged. Serum FSH and LH fall to the lower limits of detectability and are unresponsive to GnRH stimulation during pregnancy. GH concentrations are not significantly different from nonpregnant levels, but pituitary response to provocative testing is markedly altered. GH response

to hypoglycemia and arginine infusion is enhanced in early pregnancy but thereafter becomes depressed. Established pregnancy can continue in the face of hypophysectomy, and in women hypophysectomized prior to pregnancy, induction of ovulation and normal pregnancy can be achieved with appropriate gonadotropin replacement therapy. In cases of primary pituitary hyperfunction, the fetus is not affected.

Maternal Thyroid Gland

The thyroid becomes palpably enlarged during the first trimester, and a subtle bruit may be present. Thyroid iodide clearance and thyroidal ^{131}I uptake, which is clinically contraindicated in pregnancy, have been shown to be increased. These changes are due in large part to the increased renal clearance of iodide, which causes a relative iodine deficiency. Although total serum thyroxine is elevated as a result of estrogen-stimulated increased thyroid hormone–binding globulin (TBG), free thyroxine and triiodothyronine are normal (see Figure 16–2). High circulating concentrations of hCG, particularly asialo-hCG, which has weak TSH-like activity, contribute to the thyrotropic action of the placenta in early pregnancy. In fact, there is significant, though transient, biochemical hyperthyroidism associated with hCG stimulation in early gestation.

Maternal Parathyroid Gland

The net calcium requirement imposed by fetal skeletal development is estimated to be about 30 g by term. This is met by hyperplasia of the maternal parathyroid glands and elevated serum levels of parathyroid hormone. The maternal serum calcium concentration declines to a nadir at 28 to 32 weeks, largely due to the mild hypoalbuminemia of pregnancy but also to fetal bone formation. Ionized calcium is maintained at normal concentrations throughout pregnancy.

Maternal Pancreas

The nutritional demands of the fetus require alteration of maternal metabolic homeostatic control, which results in both structural and functional changes in the maternal pancreas. The size of pancreatic islets increases, and insulin-secreting beta cells undergo hyperplasia. Basal levels of insulin are lower or unchanged in early pregnancy but increase during the second trimester as a result of increased secretion rather than decreased metabolic clearance. Thereafter, pregnancy is a hyperinsulinemic state, with resistance to its peripheral metabolic effects. Pancreatic production of glucagon remains responsive to the usual endocrine stimuli and is suppressed by glucose loading, although the degree of responsiveness has not been well evaluated in pregnancy.

The major roles of insulin and glucagon involve the intracellular transport of nutrients, specifically glucose, amino acids, and fatty acids. Insulin is not transported across the placenta, but rather exerts its effects on transportable metabolites. During pregnancy, peak insulin secretion in response to meals is accelerated, and glucose tolerance curves are characteristically altered. Fasting glucose levels are maintained at low-normal levels. Excess

SYSTEM	PARAMETER	PATTERN
Cardiovascular	Heart rate	Gradually increases 20%
	Blood pressure	Gradually decreases 10% by 34 wk, then increases to prepregnancy values
	Stroke volume	Increases to maximum at 19 wk, then plateaus
	Cardiac output	Rises rapidly by 20%, then gradually increases an additional 10% by 28 wk
	Peripheral venous distention	Progressive increase to term
	Peripheral vascular resistance	Progressive decrease to term
Pulmonary	Respiratory rate	Unchanged
	Tidal volume	Increases by 30%-40%
	Expiratory reserve	Gradual decrease
	Vital capacity	Unchanged
	Respiratory minute volume	Increases by 40%
Blood	Volume	Increases by 50% in second trimester
	Hematocrit	Decreases slightly
	Fibrinogen	Increases
	Electrolytes	Unchanged
Gastrointestinal	Sphincter tone	Decreases
	Gastric emptying time	Increases
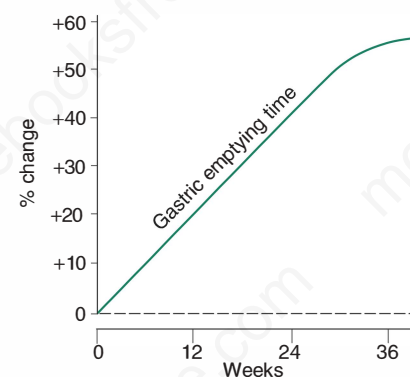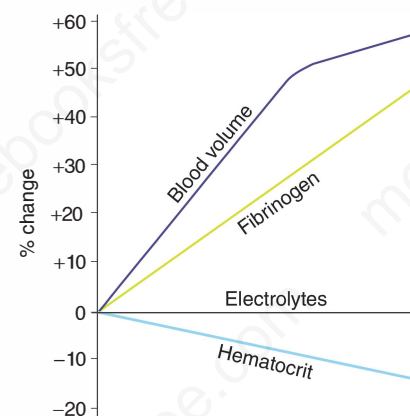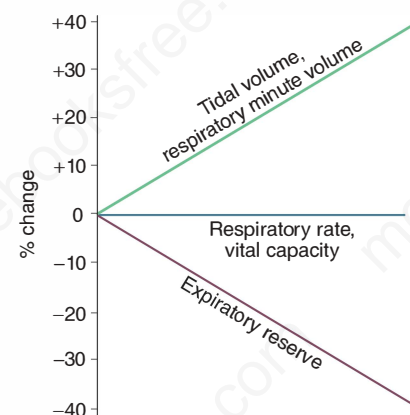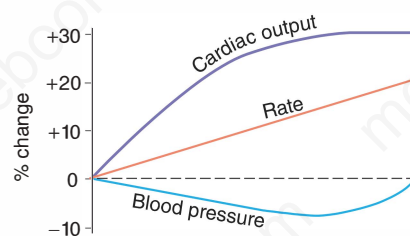
FIGURE 16–3 Maternal physiologic changes during pregnancy.

SYSTEM	PARAMETER	PATTERN
Renal	Renal flow	Increases 25%-50%
	Glomular filtration rate	Increases early, then plateaus
Weight	Uterine weight	Increases from about 60-70 g to about 900-1200 g
	Body weight	Average 11-kg (25-lb) increase
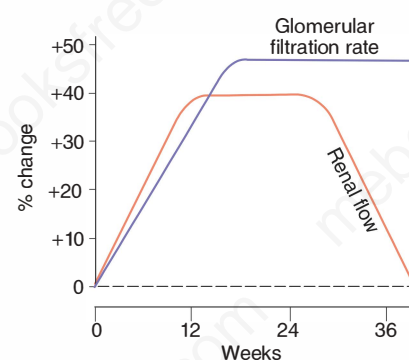
FIGURE 16–3 (*Continued*)

carbohydrate is converted to fat, and fat is readily mobilized during decreased caloric intake.

Maternal Adrenal Cortex

Total plasma cortisol concentrations increase to three times nonpregnant levels by the third trimester. Most of the changes can be explained by increased estrogens causing higher corticosteroid-binding globulin (CBG) production. The actual production of cortisol by the zona fasciculata also is increased in pregnancy. The net effect of these changes increases plasma-free cortisol, which is approximately doubled by late pregnancy. In spite of cortisol

TABLE 16–1 Effect of pregnancy on endocrine function tests.

	Test	Result
Pituitary FSH, LH	GnRH stimulation	Unresponsive from third gestational week until puerperium
GH	Insulin tolerance test	Response increases during the first half of pregnancy and then is blunted until the puerperium
	Arginine stimulation	Hyperstimulation during the first and second trimesters, then suppression
TSH	TRH stimulation	Response unchanged
Pancreas Insulin	Glucose tolerance	Peak glucose increases, and glucose concentration remains elevated longer
	Glucose challenge	Insulin concentration increases to higher peak levels
	Arginine infusion	Insulin response is blunted in mid to late pregnancy
Adrenal Cortisol	ACTH infusion	Exaggerated cortisol and aldosterone responses
	Metyrapone	Diminished response
Mineralocorticoids	ACTH infusion	No DOC response
	Dexamethasone suppression	No DOC response

concentrations approaching those found in Cushing syndrome, diurnal variation in plasma cortisol is maintained. The elevated free cortisol concentration probably contributes to the insulin resistance of pregnancy and possibly to the appearance of striae, but most signs of hypercortisolism do not occur in pregnancy. It is suggested that high progesterone levels act to antagonize glucocorticoid effects.

Serum aldosterone also is markedly elevated in pregnancy, due to an eightfold increased production rate by the zona glomerulosa. Renin substrate is increased due to the influence of estrogen on hepatic synthesis, and renin also is elevated, leading to increased angiotensin. In spite of these dramatic changes, normal pregnant women show few signs of hyperaldosteronism. There is no tendency toward hypokalemia or hypernatremia, and blood pressure at midpregnancy—when changes in the renin-angiotensin-aldosterone system are maximal—actually tends to be lower than in the nonpregnant state. Progesterone is an effective competitive inhibitor of mineralocorticoids in the distal renal tubules. Thus, increases in renin and aldosterone may simply be an appropriate response to the high gestational levels of progesterone. Elevated angiotensin II does not normally result in hypertension, because of diminished sensitivity of the maternal vascular system to angiotensin and possibly due to the vasodilatory activity of the Ang 1-7 proteolytic product. Exogenous angiotensin provokes a smaller increase in blood pressure than in the nonpregnant state. Finally, in patients with preeclampsia, the most common form of pregnancy-related hypertension, serum renin, aldosterone, and angiotensin levels are unchanged or even lower than in normal pregnancy, thus ruling out a primary role for the renin-angiotensin system in this disorder.

In normal pregnancy, the maternal production of androgens is slightly increased but these are buffered by sex hormone–binding globulin (SHBG). Levels of SHBG are increased by estrogens. Testosterone, which binds avidly to SHBG, increases to the normal male range by the end of the first trimester, but free testosterone levels are actually lower than in the nonpregnant state. DHEA sulfate does not bind significantly to SHBG, and plasma concentrations of DHEA sulfate actually decrease during pregnancy.

The desulfation of DHEA sulfate by the placenta and the conversion of DHEA to estrogens by the fetal-placental unit are important factors in its increased metabolic clearance.

FETAL ENDOCRINOLOGY

Because of the physical inaccessibility of the human fetus, much of our information about fetal endocrinology is derived indirectly. Most early studies of fetal endocrinology relied on observations of infants with congenital disorders or inferences from ablation studies or acute manipulation in experimental animals. The development of effective cell culture methods and sensitive immunoassays, as well as the ability to achieve stable preparations of chronically catheterized monkey fetuses, have increased our understanding of the dynamics of intrauterine endocrine events. The endocrine system is among the first to develop in fetal life, but its study is complicated by its relative physical isolation as well as the multiplicity of sources of the various hormones. Dating of events in fetal development is usually given in *fetal weeks*, which begin at the time of fertilization. Thus, fetal age is always 2 weeks less than gestational age.

Fetal Pituitary Hormones

The characteristic anterior pituitary cell types are discernible as early as 8 to 10 fetal weeks, and all of the hormones of the adult anterior pituitary are extractable from the fetal adenohypophysis by 12 weeks. Similarly, the hypothalamic hormones thyrotropin-releasing hormone (TRH), gonadotropin-releasing hormone (GnRH), and somatostatin are present by 8 to 10 weeks. The direct circulatory connection between hypothalamus and pituitary develops later, with capillary invasion initially visible at about 16 fetal weeks.

The role of the fetal pituitary in organogenesis during the first trimester appears to be negligible. None of the pituitary hormones are released into the fetal circulation in large quantities until after 20 fetal weeks. Even GH appears not to be influential, and in fact total absence of GH is consistent with normal development at birth. Development of the gonads and adrenals during the first trimester appears to be directed by hCG rather than by fetal pituitary hormones, so that these organs initially develop even in the face of anencephaly.

During the second trimester, there is a marked increase in secretion of all of the anterior pituitary hormones, which coincides with maturation of the hypophysial portal system. Female fetuses exhibit higher FSH levels in both pituitary and serum than do males. Differentiation of the gonads is crucial for normal male sexual development and reproductive potential in both sexes. ACTH rises significantly during the second trimester and assumes an increasing role in directing the maturation of the differentiated adrenal, as shown by the anencephalic fetus, in which the fetal zone of the adrenal undergoes atrophy after 20 weeks. Fetal PRL secretion also increases after the twentieth fetal week, but its functional significance is unknown. Vasopressin and oxytocin are demonstrable by 12 to 18 weeks in the fetal posterior pituitary gland and correlate with the development of their sites of production, the supraoptic and paraventricular nuclei, respectively.

Fetal Thyroid Gland

The fetal thyroid gland develops initially in the absence of detectable TSH. By 12 weeks the thyroid is capable of iodine-concentrating activity and thyroid hormone synthesis, but prior to 16 to 20 weeks, the maternal thyroid appears to be the primary source for fetal T_4. This has implications in that infants born to even mildly hypothyroid mothers are reported to have lower IQs and other health problems. The function of the fetal thyroid hormones appears crucial to somatic growth and for successful neonatal adaptation. In particular, auditory maturational events are regulated by thyroid hormones.

During the second trimester, TRH, TSH, and free T_4 all begin to rise. The maturation of feedback mechanisms is suggested by the subsequent plateau of TSH at about 20 fetal weeks. Fetal T_3 and reverse T_3 do not become detectable until the third trimester. The hormone produced in largest amount throughout fetal life is T_4, with the metabolically active T_3 and its inactive derivative, reverse T_3, rising in parallel to T_4 during the third trimester. At birth, conversion of T_4 to T_3 is demonstrable. Goitrogenic agents such as propylthiouracil are transferred across the placenta and may induce fetal hypothyroidism and goiter.

Fetal Adrenal Cortex

The fetal adrenal cortex is identifiable as early as 4 weeks of fetal age, and by the seventh week, steroidogenic activity can be detected in the inner zone layers.

During gestation, the adrenal cortex differs anatomically and functionally from the adult gland and occupies as much as 0.5% of total body volume. Most of this tissue is composed of a unique fetal zone that subsequently is transformed into the definitive (adult) zone during the early neonatal period. The inner fetal zone is responsible for the majority of steroids produced during fetal life and comprises 80% of the mass of the adrenal. During the second trimester, the inner fetal zone continues to grow, while the outer zone remains relatively undifferentiated. At about 25 weeks, the definitive (adult) zone develops, ultimately assuming the principal role in steroid synthesis during the early postnatal weeks.

Fetal Gonads

The testis is histologically identifiable by about 6 fetal weeks. Primary testis differentiation begins with development of the Sertoli cells at 8 weeks of gestation. SRY, the sex-determining locus on the Y chromosome directs the differentiation of the Sertoli cells, the sites of anti-Müllerian hormone (AMH) synthesis. AMH, a member of the transforming growth factor-beta family of growth factors, specifically triggers the ipsilateral resorption of the Müllerian tract in males and prevents development of female internal structures. Embryonic androgen production begins in the developing Leydig cells at about 10 weeks, coincident with the peak production of placental hCG. Binding of hCG to fetal testes with stimulation of testosterone release has been demonstrated in the laboratory. Other fetal testicular products of importance are inhibin and the 5α-reduced testosterone metabolite dihydrotestosterone. Dihydrotestosterone is responsible for development of the male external genital structures.

Little is known about early fetal ovarian function, but by 7 to 8 weeks of intrauterine life the ovaries become morphologically recognizable. Oogonia mitosis is active and steroid-producing theca cell precursors are identifiable at 20 weeks. This corresponds

with peak gonadotropin levels from the fetal pituitary. Activin and inhibin peptide subunits are expressed in the midtrimester human testis but not in the midtrimester human ovary. In contrast to the male fetus, ovarian steroid production by the fetus is not essential for female phenotypic development (see Chapter 13).

ENDOCRINE CONTROL OF PARTURITION

During the last few weeks of pregnancy, two processes herald impending labor. Sporadic uterine contractions, usually painless, become increasingly frequent, and the lower uterine segment and cervix become softer and thinner, a process known as effacement, or ripening. While false alarms are fairly common, especially at term, the onset of true labor is usually fairly abrupt, with the establishment of regular contractions every 2 to 5 minutes, leading to delivery in less than 24 hours. The exact mechanisms, however, leading to this state of parturition remain largely unknown. Reasons for this deficit of knowledge are manifold. Firstly, control of parturition in most other species is substantially different from that in humans, rendering the study of animal models for human parturition ineffective. Secondly, ethical issues largely prevent experimental studies in human pregnancy. Hence, much of the available information on human parturition has been extrapolated from animal models that may not accurately reflect the complex, species-specific process of human labor. It is assumed that parturition is a multifactorial process in which the fetus and mother mutually "cooperate." At a minimum, the placenta and amnion play key roles.

Progesterone and Nuclear Progesterone Receptors

The "progestational" effects of this steroid hormone have been elucidated since the "progesterone block" hypothesis was put forth by Arpad Csapo in the 1950s. He proposed that progesterone maintains uterine relaxation by inhibiting myometrial contractility, and that withdrawal of progesterone triggers parturition. In species where the corpus luteum is the primary source of progesterone in pregnancy (such as mice, rats, and rabbits), the demise of this ovarian structure shortly preceding parturition is mediated by prostaglandin F2α (PGF2α). Similarly in humans, up to the seventh week of pregnancy when the luteoplacental shift in progesterone synthesis occurs, luteectomy without progesterone supplementation results in abortion. Agents that inhibit the actions of progesterone, such as mifepristone (RU486), or block the conversion of pregnenolone to progesterone, induce labor in most species. While parturition is triggered by a systemic decline in plasma progesterone levels in many mammals, human parturition occurs without such a demonstrable decline. This apparent paradox has spawned the concept of a "functional progesterone withdrawal" in human labor.

Progesterone's classical actions are exerted via binding to ligand-activated nuclear receptors that function as transcription factors, modulating downstream genomic activity. In the human,

two primary progesterone receptors have been identified. Progesterone receptor A (PGR-A) and progesterone receptor B (PGR-B) are products of the same gene but under control of two distinct promoters, resulting in full-length PGR-B and truncated PGR-A isoforms. Most genomic actions of progesterone are mediated through PGR-B, whereas PGR-A primarily functions to repress PGR-B activity. Progesterone responsiveness is determined by the relative amounts of the opposing isoforms, known as the PGR-A:PGR-B ratio. Early and mid-gestation are dominated by PGR-B abundance, while in late gestation, the ratio is reversed, with a preponderance of PGR-A. Inflammation and altered processing of PGRs have been implicated in the ratio switch, resulting in "functional progesterone withdrawal" and labor.

Progesterone has been shown to directly inhibit myometrial contractility via hyperpolarization, decreased amplitude of action potentials and by augmenting activity of the cAMP/protein kinase-A (PKA) signaling cascade, promoting uterine relaxation by preventing intracellular calcium release. Progesterone and PGR-B also maintain myometrial quiescence by repression of contraction associated protein (CAP) genes. Notable CAPs include the oxytocin receptor (OTR), prostaglandin F2α receptor, various prostaglandin metabolizing enzymes and the gap junction protein, connexin 43 (Cx43). When systemic progesterone levels fall or PGR antagonists are introduced, CAPs are activated.

Estrogens and Nuclear Estrogen Receptors

Placental production of estrogens peaks before parturition (see Figure 16–2) and these hormones also play a role in the transition from uterine relaxation to contractility. Administration of estrogen boluses to non-laboring women at term increases uterine contractility within 4–6 hours and accelerates delivery. Furthermore, treatment with high doses of estrogen or androstenedione (which is converted to estrogen) at mid-gestation has been shown to precipitate preterm labor. Resistance of the myometrium to the pro-labor effects of estrogen in early gestation is likely attributable to the relatively low tissue levels of ER-α and ER-β at this stage. With the onset of parturition and change in the PGR-A/PGR-B ratio, levels of ERs dramatically increase, facilitating estrogen action. Treatment with RU486 at mid-gestation also increases ER-α levels, functionally linking progesterone and estrogen action in the pregnant myometrium.

Corticotropin-releasing Hormone (CRH)

CRH is synthesized in the placenta as well as the hypothalamus, but its regulation is distinctive. Whereas cortisol inhibits hypothalamic CRH release, in the placenta cortisol exerts positive feedback on CRH expression. Thus, within the pregnant uterus, placental CRH drives an increase in fetal pituitary ACTH, which in turn increases adrenal production of cortisol, and the latter stimulates further placental CRH production. This autoregulation is thought to be responsible for the exponential rise of maternal CRH during labor. Moreover, as term approaches there is fall in CRH binding protein levels, leading to elevated free CRH levels. Progesterone decreases CRH production, whereas cytokines, such

as TNF-α, and lipopolysaccharide stimulate CRH, providing a clinically recognized link between inflammation and labor.

Placental CRH and cortisol also act on the fetal lungs to produce surfactant protein A (SP-A), which activates NF-κB and inflammatory cytokines within the uterus, enhancing myometrial contractility and coordinating fetal pulmonary maturity with uterine activity. CRH also directly influences CAPs, such as Cx43 and the myometrial response to PGF2α.

Oxytocin

Oxytocin infusion is commonly used to induce or augment labor. Both maternal and fetal oxytocin levels increase spontaneously during labor, but neither has been convincingly shown to trigger labor, and even women with posterior pituitary failure, manifested as diabetes insipidus, are able to deliver without oxytocin augmentation. The ability of oxytocin deficient mice (OT$^{-/-}$) to deliver normally suggests that alternate pathways for labor must exist. At term, increased sensitivity of uterine tissue to oxytocin is attributed to increased synthesis and density of oxytocin receptors (OTR). Clinical studies using the OTR inhibitor, atosiban, have demonstrated that delivery can be delayed by 24 to 48 hours, but meta-analyses do not show much clinical improvement in neonatal outcome.

Prostaglandins (PGs)

PGF2α administered intra-amniotically, vaginally, or intravenously is an effective abortifacient as early as 14 weeks of gestation. PGE2 administered vaginally induces labor in most women in the third trimester. The amnion and chorion contain high concentrations of arachidonic acid, and the decidua contains active PG synthetases. As signaling molecules, PGs act in a paracrine or autocrine manner to activate intracellular signaling and gene transcription via binding to G-protein receptors. PGs are almost certainly involved in maintenance of labor once it is established and are believed to be the final common pathway of parturition.

PRETERM LABOR/BIRTH

Preterm birth (PTB) is one of the most important problems in obstetrics today as it complicates approximately 12% of all births in the United States and 5% to 9% in other developed countries. Complications of preterm birth are significant; prematurity accounts for 75% of perinatal mortalities and more than 50% of long-term infant morbidities, resulting in high health-care costs. Approximately 50% of PTB is idiopathic or spontaneous, leading to delivery before 37 weeks of gestation. Thirty percent of cases follow preterm rupture of membranes (PROM) and the other 20% are attributed to medically indicated preterm delivery, typically due to worsening maternal disease or fetal compromise. The PTB rate rose more than 30% from 1981 to 2006, with late preterm births (at 34-36 weeks of gestational age) accounting for two-thirds of the increase. Annual costs of PTB have been calculated to approach 10 billion dollars in the USA alone.

PTB has a complex pathophysiology with heterogeneous genetic and environmental susceptibilities. Risk factors for preterm labor are varied and include history of prior preterm labor, uterine anomalies, multiple gestation, maternal medical complications, low prepregnancy body mass index, gestational bleeding, low socioeconomic status, minority race, behavioral habits such as smoking, alcohol and drug abuse, stressful events, limited or no prenatal care, and infection. The etiology of preterm labor is not well understood but may result from mechanical factors, hormonal changes, infection, or interruption of normal mechanisms responsible for sustaining uterine quiescence. Survival rates have improved due to the use of antenatal corticosteroids and improvements in neonatal resuscitation and intensive care. There is a well-documented racial disparity in rates of PTB with African Americans at 17% versus only 11% in Caucasians. Further, the infant mortality rate for African-American babies is 2.5 times higher than Caucasians. Population-based studies indicate that this disparity cannot be explained by socioeconomic status, maternal behavior, marital status, or education alone and likely point to true genetic differences or social stress.

Four major mechanistic pathways leading to PTB have been proposed: (1) stress or activation of maternal or fetal hypothalamic-pituitary-adrenal axis, (2) pathologic uterine distention, (3) inflammation/infection, and (4) decidual hemorrhage and coagulation. These are summarized in Figure 16–4. Each of these pathways has different epidemiologic, clinical, and biochemical features, but all are likely to initiate labor via PG production, myometrial activation, and degradation of extracellular matrix components leading to premature rupture of membranes and cervical dilation. The fetoplacental endocrine unit can trigger labor prematurely, activating CRH release in cases of a hostile intrauterine environment such as stress or chorioamnionitis.

Endotoxins from bacteria induce proinflammatory mediators such as IL-1β and TNF-α from the decidua, which stimulate PG production. Double knockout mice missing IL-1β and TNF-α have decreased rates of PTB after administration of microorganisms. Inflammatory cytokines in uterine tissue have been shown to increase the stability of PGR-A and affect the PGR-A:PGR-B ratio, precipitating functional progesterone withdrawal. Furthermore cytokines may partly explain racial disparity in PTB. Elevated IL-1β and TNF-α concentrations are seen in African-American cases of PTB, whereas high IL-6 and IL-8 levels are more commonly associated with preterm birth in Caucasians. Single nucleotide polymorphisms in genes regulating or encoding these cytokines may be a cause of this disparity.

Predictors/Prevention of Preterm Labor

Unfortunately, clinical symptoms are too inaccurate and unreliable to identify women at risk for PTB. Given the medical, emotional, and financial gravity of preterm birth, many tests have been promoted to identify women at risk for PTB. Home uterine activity monitors are not recommended as they have not been shown to improve outcome. The presence of bacterial vaginosis (BV) has been associated with preterm delivery independent of other known risk factors, however, insufficient data support screening
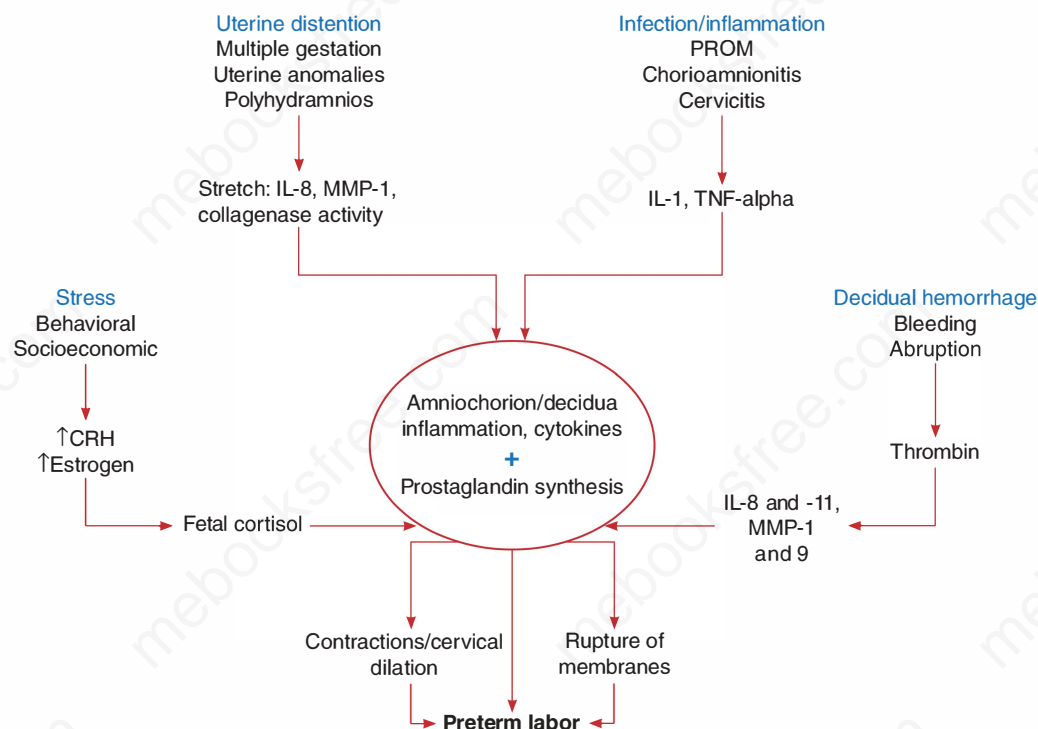
FIGURE 16–4 Major mechanistic pathways leading to preterm birth.

and treating women with positive cultures. Maternal estriol levels have been used to predict activation of the fetal hypothalamic-pituitary-adrenal axis, but trials failed to establish accurate prediction of preterm labor. Numerous studies have confirmed the association between cervical shortening and subsequent preterm delivery. Transvaginal cervical ultrasonography has been shown to be a reliable way to assess the length of the cervix and is a useful screen in women with a history of PTB. Studies have shown that supplementation of vaginal progesterone in women without prior PTB and with a cervical length <2.5 cm reduces the risk of PTB. In patients with a cervix shorter than 2.5 cm, cerclage is an effective strategy. Weekly injections of 17α-hydroxyprogesterone caproate from 16 to 20 weeks until 36 weeks of gestation have been found to significantly reduce recurrent PTB in women with a past history of preterm delivery. A meta-analysis of 10 placebo-controlled trials found progesterone therapy reduced the incidence of recurrent preterm birth from 36% to 26%. Given the lack of other available treatments, many providers are using progesterone as preventative therapy in high-risk patients, but progesterone does not reverse established labor. Fetal fibronectin (fFN) is a basement membrane protein of placental membranes and decidua. Clinical trials have shown that the risk of PTB is increased in the presence of fFN and decreased in its absence. A negative fFN test indicates a 95% likelihood of not delivering within 14 days; therefore it is particularly useful for ruling-out preterm delivery.

Management of Preterm Labor

Once preterm labor has been diagnosed, care should be taken to determine if patient warrants tocolysis, expectant management, or

delivery. It is critical to confirm an accurate gestational age because if less than 34 weeks, glucocorticoids should be administered to promote neonatal lung maturity. Tocolytic therapy is used to attempt to prolong pregnancy to allow time for corticosteroid administration. In general, tocolytic drugs prolong pregnancy for 2 to 7 days. There is no first-line tocolytic drug and the choices include beta-mimetics, calcium channel blockers, and NSAIDs. In Europe, the OTR antagonist atosiban is approved. Tocolytic choice should be based on maternal/fetal conditions, side effects, and gestational age. Serious adverse events are rare but potentially life-threatening. Beta-mimetics and calcium channel blockers are associated with a risk of pulmonary edema. Magnesium sulfate may cause maternal lethargy, nausea and vomiting and respiratory depression and is no longer recommended. NSAIDs can cause fetal oligohydramnios, premature closure of the ductus arteriosus and neonatal necrotizing enterocolitis. Calcium channel blockers used as a single agent appear to have a good maternal and fetal safety profile. Combining tocolytic drugs potentially increases maternal morbidity and should be used with caution.

POSTTERM PREGNANCY

Postterm pregnancy refers to gestation that has extended to or beyond 42 weeks and is associated with significant risks to the fetus. Most importantly, the perinatal mortality rate (stillbirths plus early neonatal deaths) beyond 42 weeks is twice that at term (4-7 deaths vs 2-3 deaths per 1000 deliveries) and is sixfold higher after 43 weeks or more. Potential causes include uteroplacental insufficiency, meconium aspiration, and intrauterine infection.

Postterm pregnancy is also a risk factor for low Apgar scores, oligohydramnios, and neonatal complications such as hypoglycemia, seizures, respiratory insufficiency, and increased risk of death within the first year of life. Postterm pregnancy is associated with some complications for the mother including a twofold increased risk of severe perineal injury and cesarean delivery. Because pregnancy past 41 weeks is associated with high costs of antenatal testing and induction of labor, as well as a source of significant anxiety for pregnant women, there has been a movement in obstetrics to effect delivery before 41 completed weeks. The most common cause of a prolonged gestation is incorrect dating; other risk factors include primiparity, prior postterm pregnancy, and male sex of the fetus. Recently, obesity has been identified as a modifiable risk factor. The mechanism for this association is unclear but it is likely due to a derangement in endocrine factors that initiate labor. In rare cases, postterm pregnancy has been associated with placental sulfatase deficiency or fetal anencephaly.

Management of Postterm Pregnancy

Standard management of postterm pregnancy focuses on antenatal fetal surveillance and timely initiation of delivery if spontaneous labor does not occur. Although antepartum testing for postterm pregnancies is universally accepted, it has not been validated in prospective trials to decrease perinatal mortality. Given the ethical and medicolegal concerns for a control group, randomized studies to prove benefit will likely never be done. Timing of delivery is generally recommended when risks to continuing the pregnancy are greater than the benefit to mother or fetus after or during birth. In high-risk cases such as diabetics and hypertensives, pregnancy should not be allowed to progress beyond 41 weeks. However, management of low-risk postterm pregnancy is more controversial. Because vaginal delivery cannot always be induced readily, maternal risks and considerations complicate this decision. Factors to consider include gestational age, results of antepartum fetal testing, cervical examination, and maternal preference after counseling regarding the risks, benefits, and alternatives to expectant management with antepartum monitoring versus labor induction.

ENDOCRINOLOGY OF THE PUERPERIUM

Extirpation of any active endocrine organ leads to compensatory changes in other organs and systems. Delivery of the infant and placenta causes both immediate and long-term adjustment to loss of the pregnancy hormones. The sudden withdrawal of fetal-placental hormones at delivery permits determination of their serum half-lives and some evaluation of their effects on maternal systems.

Physiologic and Anatomic Changes

Some physiologic and anatomic adjustments that take place after delivery are both hormone dependent and independent. For example, major readjustments of the cardiovascular system occur in response to the normal blood losses associated with delivery and to loss of the low-resistance placental shunt. By the third postpartum day, blood volume is estimated to decline to about 80% of predelivery values. These cardiovascular changes influence renal and liver function, including clearance of pregnancy hormones.

Uterine Changes

The uterus involutes progressively at the rate of about 500 g/wk and continues to be palpable abdominally until about 2 weeks postpartum, after which it reoccupies its position entirely within the pelvis. Nonpregnant size and weight (60-70 g) are reached by 6 weeks postpartum. The reversal of myometrial hypertrophy occurs with a decrease in size of individual myometrial cells rather than by reduction in number. The endometrium, which is sloughed at the time of delivery, regenerates rapidly; by the seventh postpartum day there is restoration of surface epithelium. These rapid regenerative changes do not apply to the area of placental implantation, which requires much longer for restoration and retains pathognomonic histologic evidence of placentation indefinitely.

Endocrine Changes

A. Steroids   With expulsion of the placenta, many steroid levels decline precipitously, with half-lives of minutes. Plasma progesterone falls to luteal-phase levels within 24 hours after delivery and to follicular-phase levels within several days. Removal of the corpus luteum results in a fall to follicular levels within 24 hours. Estradiol reaches follicular phase levels within 1 to 3 days after delivery.

B. Pituitary hormones   The pituitary gland, which enlarges during pregnancy owing primarily to an increase in lactotrophs, does not diminish in size until after lactation ceases. Secretion of FSH and LH continues to be suppressed during the early weeks of the puerperium, and stimulus with bolus doses of GnRH results in subnormal release of LH and FSH. Over the ensuing weeks, responsiveness to GnRH gradually returns to normal and most women exhibit follicular-phase serum levels of LH and FSH by the third or fourth postpartum week.

C. Prolactin   Serum PRL, which rises throughout pregnancy, falls with the onset of labor and then exhibits variable patterns of secretion depending on whether breast-feeding occurs. Delivery is associated with a surge in PRL, which is followed by a rapid fall in serum concentrations over 7 to 14 days in the nonlactating mother. In nonlactating women, the return of normal cyclic function and ovulation may be expected within 3 months, with the initial ovulation occurring at an average of 9 to 10 weeks postpartum. In actively lactating women, PRL can cause a persistence of anovulation. Surges of PRL are believed to act on the hypothalamus to inhibit GnRH secretion. Administration of exogenous GnRH during this time induces normal pituitary responsiveness, and occasional ovulation may occur spontaneously even during lactation.

LACTATION

Development of the breast alveolar lobules occurs throughout pregnancy. This period of mammogenesis requires the concerted participation of estrogen, progesterone, PRL, GH, and glucocorticoids. hPL may also play a role but is not indispensable. Lactation requires PRL, insulin, and adrenal steroids and is associated with further enlargement of the lobules, followed by synthesis of milk constituents such as lactose and casein. Lactopoiesis does not occur until unconjugated estrogens fall to nonpregnant levels at about 36 to 48 hours postpartum. Evidence from GH-deficient and hypothyroid patients suggests that GH and thyroid hormone are not required.

PRL is essential to milk production. Its action involves induced synthesis of large numbers of PRL receptors. These receptors appear to be autoregulated because PRL increases receptor levels in cell culture and bromocriptine, an inhibitor of PRL release, causes a decrease in both PRL and its receptors. In the absence of PRL, milk secretion does not take place. However, even in the presence of high levels of PRL during the third trimester, milk secretion does not take place until after delivery, due to the blocking effect of high levels of estrogen.

Milk secretion requires the additional stimulus of suckling, which activates a neural arc. Milk ejection occurs in response to a surge of oxytocin, which induces a contractile response in the smooth muscle surrounding the gland ductules. Oxytocin release is occasioned by stimuli of a visual, psychologic, or physical nature that prepare the mother for breast-feeding; whereas PRL release is limited to the suckling reflex arc.

ENDOCRINE DISORDERS AND PREGNANCY

Hyperthyroidism in Pregnancy

Pregnancy mimics hyperthyroidism because it causes thyroid enlargement, increased cardiac output, and peripheral vasodilation. Owing to the increase in TBG in pregnancy, total serum thyroxine is in the range expected for hyperthyroidism. Free thyroxine, the free thyroxine index, and TSH levels, however, remain in the normal range (see Chapter 7).

True hyperthyroidism complicates 1 to 2 per 1000 pregnancies. The most common form of hyperthyroidism during pregnancy is Graves disease. Hyperthyroidism is associated with an increased risk of preterm delivery (11%-25%) and may modestly increase the risk of early abortion. In Graves disease, thyroid-stimulating immunoglobulin, a 7S immune gamma globulin, is actively transported across the placenta and may cause fetal goiter and transient neonatal hyperthyroidism.

A. Management of the mother   The treatment of maternal hyperthyroidism is complicated in pregnancy. Radioiodine is strictly contraindicated. Iodide therapy can lead to huge fetal goiter and is contraindicated except as acute therapy to prevent thyroid storm before thyroid surgery. All antithyroid drugs cross the placenta and may cause fetal hypothyroidism and goiter or cretinism in the newborn. However, propylthiouracil in doses of 300 mg/d or less has been shown to be reasonably safe, although even at low doses about 10% of newborns will have a detectable goiter. A reasonable plan of management is to begin therapy in the first trimester with propylthiouracil in doses high enough to bring the free T_4 index into the mildly hyperthyroid range and then to taper the dose gradually. Despite its risk of aplasia cutis, switching to methimazole for the second and third trimester is recommended. Propranolol has been used to control maternal cardiovascular symptoms but may result in fetal bradycardia, growth restriction, preterm labor, and neonatal respiratory depression. Partial or total thyroidectomy, especially in the second trimester, is a reasonably safe procedure except for the risk of preterm labor.

B. Management of the newborn   Newborns should be observed carefully. In infants of mothers given propylthiouracil, even equivocal evidence of hypothyroidism is an indication for thyroxine replacement therapy. Neonatal Graves disease, which may present as late as 2 weeks after delivery, requires intensive therapy (see Chapter 7).

Hypothyroidism in Pregnancy

Clinical hypothyroidism is uncommon in pregnancy because most women with the untreated disorder are anovulatory. As a practical matter, women taking thyroid medication at the time of conception should be maintained on a 30% higher dose throughout pregnancy. Physiologic doses of thyroid hormone replacement are innocuous, but maternal hypothyroidism is hazardous to the developing fetus. Women with a personal or family history of thyroid disease or with symptoms suggestive of hypothyroidism should have TSH tested prior to conception. The correlation between maternal and fetal thyroid status is poor, and hypothyroid mothers frequently deliver euthyroid infants. The strongest correlation between maternal and newborn hypothyroidism occurs in areas where endemic goiter due to iodide deficiency is common. In an iodine-deficient population, treatment with iodine in the first and second trimesters of pregnancy significantly reduces the incidence of cretinism (growth failure, mental retardation, and other neuropsychologic deficits). Even subclinical maternal hypothyroidism has been associated with fetal neurodevelopmental damage and reduced IQs in the offspring. This is thought to be due to inadequate transplacental supply of thyroxine T_4 in early pregnancy as the mother is the only source of T_4 for the developing first trimester fetal brain.

Pituitary Disorders in Pregnancy

In women of reproductive age, small tumors of the anterior pituitary are not uncommon (see also Chapter 4). Although many are nonfunctional and asymptomatic, the most common symptom of pituitary microadenomas is amenorrhea, frequently accompanied by galactorrhea. In the past, few affected women became pregnant, but now most can be made to ovulate and to conceive with the aid of clomiphene citrate, recombinant gonadotropins and

hCG, or bromocriptine. Before ovulation is induced in any patient, serum PRL should be determined. Modest elevations of PRL warrant checking IGF-I levels, because hyperprolactinemia occurs in 25% of patients with GH-producing adenomas. If PRL or IGF-1 is elevated, the sella turcica should be evaluated by magnetic resonance imaging. About 10% of women with secondary amenorrhea are found to have adenomas, and 20% to 50% of women with amenorrhea and galactorrhea have detectable tumors.

The effect of pregnancy on pituitary adenomas depends on their size. Microadenomas rarely (1%) lead to visual field defects. However, approximately 20% of macroadenomas may lead to abnormalities in visual fields or other neurologic signs, usually in the first half of pregnancy. These changes almost always revert to normal after delivery, so aggressive therapy for known pituitary adenomas is not indicated except in cases of rapidly progressive visual loss, where surgery is indicated. Monitoring of patients with known PRL-secreting adenomas during pregnancy is primarily based on clinical examination. The normal gestational increase in PRL may obscure an increase attributable to the adenoma, and radiographic procedures are undesirable in pregnancy.

A. Management

Management of the pregnant woman with a pituitary macroadenoma includes early ophthalmologic consultation for formal visual field mapping and repeat examinations every other month throughout pregnancy. If visual field disturbances are minimal, pregnancy may be allowed to proceed to term. If symptoms become progressively more severe and the fetus is mature, labor should be induced. If symptoms are severe and the fetus is immature, management may consist of transsphenoidal resection of the adenoma or medical treatment with bromocriptine. Although bromocriptine inhibits both fetal and maternal pituitary PRL secretion, it does not affect decidual PRL secretion. Bromocriptine appears not to be teratogenic, and no adverse fetal effects have been reported. In most cases, it is probably preferable to surgery. The newer, selective dopamine D_2 receptor agonist cabergoline has shown excellent results in normalizing PRL levels, inducing tumor shrinkage, and minimizing side effects. The data available on pregnancies in which cabergoline was used do not show adverse effects; however, because the data are limited compared with the large numbers for bromocriptine safety, the latter is recommended. Both have category B classification for pregnancy. Radiation therapy should not be used in pregnancy, however, focused gamma knife pituitary radiation, with abdominal shielding, is gaining popularity in this setting with less than 0.01% of the radiation dose reaching the uterus.

Lymphocytic hypophysitis is an autoimmune inflammation of the pituitary that classically occurs in women in late pregnancy or during the puerperium. The clinical presentation is difficult to distinguish from that of a large PRL-secreting adenoma. Expectant conservative medical management with corticosteroids is sometimes possible, with vigilant postpartum observation to prevent consequences of pituitary insufficiency.

B. Prognosis and follow-up

There appears to be no increase in obstetric complications associated with pituitary adenomas, and no fetal jeopardy. The rate of prematurity increases in women with tumors requiring therapy, but this is probably due to aggressive intervention rather than to direct effects on spontaneous preterm labor. The postpartum period is characterized by rapid relief of even severe symptoms, with less than 4% of untreated tumors developing permanent sequelae. In some cases, tumors improve following pregnancy, with normalization or lowering of PRL relative to prepregnancy values. Management should include imaging and assessment of PRL levels 4 to 6 weeks after delivery. There are no contraindications to breast-feeding.

Obesity and Pregnancy

Obesity is a complicated metabolic and endocrine condition that has been implicated in a number of pregnancy-related complications. Its prevalence has dramatically increased in the United States over the last 20 years. Currently approximately one-third of adult women are obese (body mass index [BMI] >30). This is a particular problem for non-Hispanic black women (49%) and Mexican American women (38%). Obese women are at increased risk for a number of adverse outcomes during pregnancy, including spontaneous abortion, stillbirth, preeclampsia, gestational diabetes, congenital anomalies, cesarean section, venous thromboembolism, and increased surgical morbidity. Complications in pregnancy are generally related to maternal pregravid obesity rather than excessive weight gain during pregnancy. In 2009, the Institute of Medicine updated its guidelines recommending weight gain of 25 to 35 lb for pregnant women with a normal BMI, 15 to 25 lb for overweight women, and 11 to 20 lb for obese women. These were initially designed to help prevent fetal growth restriction. However, most overweight/obese women have mean weight gain exceeding these guidelines. Because of the risks to the fetus, it is recommended that women attempt to lose weight prior to conception.

A. Early pregnancy risks associated with obesity

There is a well-documented increased risk of early miscarriage in obese women in both spontaneous pregnancies and after infertility treatments. Maternal obesity is also associated with an increased risk of a range of structural anomalies particularly neural tube defects, heart defects, and omphalocele. As these types of anomalies are also associated with diabetes, it has been suggested that some of these women may be undiagnosed diabetics. To complicate this further, obesity can affect accuracy of diagnostic testing for congenital anomalies. Heavier women have lower AFP levels likely from greater plasma volume and therefore adjustments have to be made for serum screening based on maternal weight. This group also is more likely to have uninterpretable cell-free fetal DNA testing due to a reduced fetal fraction DNA concentration. Ultrasound is the primary diagnostic tool for identifying structural abnormalities. Increasing maternal weight is associated with increasing impairment of adequate ultrasound visualization particularly of cardiac and craniospinal abnormalities. For these reasons, it is important to encourage preconception weight loss and perform early screening for diabetes.

B. Late pregnancy risks associated with obesity

Maternal obesity in pregnancy is associated with a significantly increased rate of gestational diabetes (GDM), preeclampsia, and stillbirths. In a large prospective study of 16,102 women, obese women and morbidly obese women were 2.5 and 3.2 more likely, respectively, to develop gestational hypertension, and 1.6 and 3.3 times more likely to develop preeclampsia. Given the potential increase in oxidative stress from maternal adipose tissue, antioxidants such as vitamin C and E have been studied in the prevention of preeclampsia in obese women; however, they have yet to show any benefit. Gestational diabetes occurs due to decreased insulin sensitivity and inadequate insulin secretory response. Obese women, in general, are more insulin resistant at baseline; therefore, it is not surprising that they are at higher risk of developing GDM. Overweight and obese women have a 2.0 to 2.4 fold increased risk of unexplained fetal death, respectively. The exact pathophysiology of unexplained intrauterine fetal death in obese women is currently unknown.

C. Peripartum risks associated with obesity

Overweight and obese women also have an increased rate of cesarean section and surgical complications, including risks of anesthesia, wound infection, excessive blood loss, endometritis, and deep venous thrombosis. In a recent report, the rate of cesarean section in overweight women was 33.8% and 47% in obese women, compared to only 20.7% in normal weight controls.

D. Long-term risks associated with obesity

Maternal obesity is a risk factor for fetal macrosomia. Accumulating evidence indicates fetal macrosomia is associated with adolescent and adult obesity and metabolic syndrome. A large retrospective study recently found that children born to obese mothers were twice as likely to be obese at 2 years of age. Both maternal obesity and maternal diabetes may independently affect the risk of adolescent obesity in children. Macrosomic infants are also at increased risk of subsequent development of diabetes. Thus, the epidemics of obesity and diabetes may continue to increase further as a result of fetal overgrowth and adiposity in utero.

Parathyroid Disease and Pregnancy

Pregnancy is associated with an approximately 25 to 30 g accumulation of calcium in the mother by term. Most of this calcium is used for fetal skeletal development. Elevated levels of 1,25-hydroxyvitamin D from the decidua leads to increased calcium absorption by the maternal gastrointestinal tract. The kidneys also work to conserve calcium during pregnancy by increasing reabsorption. Overall PTH levels are slightly lower throughout pregnancy and calcium is actively transported across the placenta to the fetus against a concentration gradient. The fetus is relatively hypercalcemic, hypercalcitonemic, and hypoparathyroid in comparison to the mother, but this resolves shortly after birth.

A. Hyperparathyroidism

Hyperparathyroidism can be either primary, from an overactive parathyroid gland(s), or a secondary physiologic response to low calcium or vitamin D deficiency. Primary hyperparathyroidism in pregnancy is rare and is due to a solitary or multiple parathyroid adenomas, hyperplasia, or carcinoma. Symptoms of hyperparathyroidism are nonspecific including fatigue, anorexia, nausea, vomiting, constipation, mental status changes, and depression and are generally present when serum calcium levels are over 12 mg/dL. Interestingly, women with hyperparathyroidism typically experience improvement in their symptoms and calcium levels during pregnancy, due to the large amount of calcium taken up by the placenta for the fetus. Complications during pregnancy from uncontrolled hyperparathyroidism include nephrolithiasis, pancreatitis, hyperemesis gravidarum, and hypercalcemic crisis. There is an increased incidence of these complications during the postpartum period when the fetal/placental drain of calcium has been removed if the hyperparathyroidism persists after birth of the infant. Fetal complications have been reported to be as high as 80% and include increased risk of spontaneous abortions, intrauterine growth restriction, low birth weight, intrauterine demise, and postpartum neonatal hypocalcemia/tetany due to suppression of the fetal parathyroid glands. In general, neonates do well if supplemental calcium and vitamin D are started promptly and serum biochemistries are closely monitored. If possible, delivery should be delayed until eucalcemia is achieved in order to allow for normal development of the fetal parathyroid glands. Surgical excision of a symptomatic adenoma is generally recommended in pregnancy during the second trimester. Conservative management throughout pregnancy is, however, possible with close maternal and fetal surveillance, if the patient is asymptomatic and only mild elevations in serum calcium are present.

B. Hypoparathyroidism

Hypoparathyroidism in pregnancy is even rarer than hyperparathyroidism and is most commonly due to incidental resection of or damage to the parathyroid glands during thyroidectomy. Symptoms from hypocalcemia include numbness and tingling of fingers and perioral area. If serum calcium levels are maintained within normal limits, pregnancy outcome should not be negatively affected, but large amounts may be needed. If not treated, maternal hypocalcemia can lead to hyperparathyroidism in the fetus causing bone demineralization, and labor may be complicated by tetany. Treatment of hypoparathyroidism involves administration of supplemental calcium and vitamin D or its active metabolite 1,25-dihydroxyvitamin D.

PREECLAMPSIA/ECLAMPSIA

Preeclampsia is an idiopathic multisystem disease seen only in pregnancy and during the peripartum period. The incidence of preeclampsia is approximately 7%. Risk factors include nulliparity, African-American race, extremes of age, chronic hypertension, multiple gestation, prior history of preeclampsia, family history of preeclampsia, chronic renal disease, antiphospholipid antibody syndrome, hydramnios, diabetes, and obesity. Preeclampsia is characterized by onset of hypertension with proteinuria or end-organ dysfunction after 20 weeks of gestation. Other signs and symptoms can include proteinuria, headache, visual disturbances,

epigastric pain, nausea/vomiting, thrombocytopenia, abnormal liver and kidney function tests. Eclampsia is the occurrence of seizures that cannot be attributed to other causes in women with preeclampsia and is associated with a maternal mortality rate as high as 10%. Preeclampsia accounts for approximately 15% of maternal deaths in the United States. Deaths occur most frequently from cerebral hemorrhage, renal failure, disseminated intravascular coagulopathy, acute pulmonary edema, or hepatic failure. Preeclampsia is also associated with perinatal fetal death and morbidity, due most commonly to iatrogenic prematurity.

Pathophysiology

The exact pathogenesis of preeclampsia remains poorly understood despite a significant amount of research in the field. It likely involves both maternal and placental factors as it is also seen in molar pregnancies with only trophoblastic and no fetal tissue. Abnormalities in placental vasculature development are postulated to result in placental hypoxia, which leads to factors being released that cause hypertension and microangiopathy of target organs. Both epidemiologic and experimental data support the role of the placenta in the etiology of preeclampsia as placental tissue is necessary for the development of the disease and once the placenta is removed its signs and symptoms begin to abate.

In preeclampsia, cytotrophoblast invasion is shallow and only the decidual portion of the spiral arteries is remodeled. In contrast, in normal pregnancies the cytotrophoblasts infiltrate deeper into the muscular tunica media of the spiral arteries, transforming these small muscular arterioles into large vessels with low resistance that facilitate placental blow flow (see Figure 16–1). It is likely that environmental, immunologic, and genetic factors all play a role in this important process.

Hypoperfusion and hypoxia are likely both a cause and effect of abnormal placental vascularization and subsequent preeclampsia. Animal models support this theory as induced preeclampsia generally requires mechanically reduced uteroplacental perfusion. In addition, maternal medical conditions associated with vascular dysfunction such as hypertension, lupus, and renal disease increase the risk of preeclampsia, as do obstetric conditions that increase placental volume without increasing placental blood flow, such as twins and hydatidiform moles.

Generalized endothelial dysfunction may explain many of the multiple clinical features of preeclampsia. Evidence to support this includes enhanced vascular reactivity to angiotensin II, impaired flow-mediated vasodilation, decreased production of endothelial-derived vasodilators such as nitric oxide and prostacyclin, and increased production of vasoconstrictors like endothelins and thromboxanes.

Extensive placental angiogenesis is required to supply the necessary nutrients and oxygen to the fetus. Normal placental vessel development relies on a balance between proangiogenic factors such as VEGF and placental growth factor (PlGF) and antiangiogenic factors such as soluble fms-like tyrosine kinase 1 (sFlt-1) and soluble endoglin (sEng). It is proposed that free sFlt-1 in the maternal circulation antagonizes the angiogenic activities of VEGF and PlGF, leading to ischemia and vasospasm. sEng from

the placenta is elevated in preeclamptic women, and appears to inhibit TGF-beta-1-mediated vasodilation. A recent study indicates that a low ratio of sFlt-1:PlGF has a negative predictive value of greater than 99% for the development of preeclampsia in the ensuing week.

Despite extensive research and multiple hypotheses regarding its etiology and pathophysiology, there is still no unifying theory that explains all the features of preeclampsia. It is generally accepted that the basis for the development of preeclampsia occurs early in pregnancy.

Clinical Features

Preeclampsia is classified as having severe features if any of the following signs or symptoms are present: headache, blurred vision, scotomata, altered mental status, right upper quadrant or epigastric pain, severe hypertension (160/110 mm Hg), pulmonary edema, oliguria (<500 cc/24 h), renal failure (creatinine >1 mg/dL over baseline), elevated liver function tests (>2× normal), thrombocytopenia (platelets <100,000), coagulopathy or HELLP syndrome (hemolysis, elevated liver enzymes, low platelets). Despite knowledge of the many risk factors described above, it is still not possible to determine which mothers will ultimately develop preeclampsia and there are no effective preventative treatments. Large clinical trials using calcium and antioxidant vitamin supplementation revealed minimal to no benefit in preventing the development of preeclampsia. Although some clinical trials of low dose (81 mg/d) aspirin failed to show benefit, several meta-analyses support its use for prophylaxis.

Treatment/Management of Preeclampsia

The only definitive therapy for preeclampsia is delivery of fetus and placenta. This is straightforward for pregnancies near term. However, pregnancies between 24 and 34 weeks are more complicated as the obstetrician must balance the significant perinatal morbidity and mortality associated with preterm delivery versus the maternal and fetal risk of prolonging these high-risk pregnancies. Antihypertensive medications should be used in setting of severe hypertension to prevent a maternal stroke; however, their use in mild to moderate hypertension has not been shown to improve maternal or fetal outcome. Magnesium sulfate therapy is initiated as the anticonvulsant of choice for preeclamptics during labor, while administering corticosteroids prior to a planned preterm delivery, and for 24 hours postpartum.

REFERENCES

General

Edlow AG, Norwitz ER. Endocrine diseases of pregnancy. In: Strauss JF, Barbieri RL, eds. *Yen and Jaffe's Reproductive Endocrinology. Physiology, Pathophysiology and Clinical Management.* 7th ed. Philadelphia: Elsevier Saunders; 2014: 604-650.

Placental Steroid Hormones

Pařízek A, Koucký M, Dušková M. Progesterone, inflammation and preterm labor. *J Steroid Biochem Mol Biol*. 2014;139:159. [PMID: 23454115]

Romero R, Stanczyk FZ. Progesterone is not the same as 17α-hydroxyprogesterone caproate: implications for obstetrical practice. *Am J Obstet Gynecol*. 2013; 208(6):421. [PMID: 23643669]

Aneuploidy Screening

Driscoll DA, Gross SJ. Screening for fetal aneuploidy and neural tube defects. Professional Practice Guidelines Committee. *Genet Med*. 2009;11:818. [PMID: 19915395]

Spencer K. Screening for Down syndrome. *Scand J Clin Lab Invest Suppl*. 2014; 244:41. [PMID: 25083892]

Fetal Endocrinology

Miller WL, Witchel SF. Prenatal treatment of congenital adrenal hyperplasia: risks outweigh benefits. *Am J Obstet Gynecol*. 2013;208:354. [PMID: 23123167]

Thyroid and Pituitary

Arrowsmith S, Wray S. Oxytocin: its mechanism of action and receptor signalling in the myometrium. *J Neuroendocrinol*. 2014;26:356. [PMID: 24888645]

Molitch ME. Endocrinology in pregnancy: management of the pregnant patient with a prolactinoma. *Eur J Endocrinol*. 2015;172:R205. [PMID: 25805896]

Yalamanchi S, Cooper DS. Thyroid disorders in pregnancy. *Curr Opin Obstet Gynecol*. 2015;27:406. [PMID: 26485458]

Obesity

Catalano P, de Mouzon SH. Maternal obesity and metabolic risk to the offspring: why lifestyle interventions may have not achieved the desired outcomes. *Int J Obes (Lond)*. 2015;39:642. [PMID: 25777180]

Gilmore LA, Klempel-Donchenko M, Redman LM. Pregnancy as a window to future health: excessive gestational weight gain and obesity. *Semin Perinatol*. 2015;39:296. [PMID: 26096078]

Parturition

Florio P, Romero R, Chaiworapongsa T, et al. Amniotic fluid and umbilical cord plasma corticotropin-releasing factor (CRF), CRF-binding protein, adrenocorticotropin, and cortisol concentrations in intraamniotic infection and inflammation at term. *J Clin Endocrinol Metab*. 2008;93:3604. [PMID: 18559919]

Mesiano S, Wang Y, Norwitz ER. Progesterone receptors in the human pregnancy uterus: do they hold the key to birth timing? *Reprod Sci*. 2011;18:6. [PMID: 20889955]

Smith R, Paul J, Maiti K, Tolosa J, Madsen G. Recent advances in understanding the endocrinology of human birth. *Trends Endocrinol Metab*. 2012;23:516. [PMID: 22682480]

Hypertension and Preeclampsia

American College of Obstetricians and Gynecologists, Task Force on Hypertension in Pregnancy. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. *Obstet Gynecol*. 2013;122:1122. [PMID: 24150027]

Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. *BMJ*. 2013;347:f6564. [PMID: 24201165]

Conrad KP, Davison JM. The renal circulation in normal pregnancy and preeclampsia: is there a place for relaxin? *Am J Physiol Renal Physiol*. 2014;306:F1121. [PMID: 24647709]

Dechend R, LaMarca B, Taylor RN. The renin-angiotensin system, its autoantibodies, and body fluid volume in preeclampsia. In: Taylor RN, Roberts JM, Cunningham FG, Lindheimer MD, eds. *Chesley's Hypertensive Disorders of Pregnancy. 4th ed*. Amsterdam: Elsevier-Academic Press; 2015:315-334.

Karumanchi SA, Rana S, Taylor RN. Angiogenesis and preeclampsia. In: Taylor RN, Roberts JM, Cunningham FG, Lindheimer MD, eds. *Chesley's Hypertensive Disorders of Pregnancy. 4th ed*. Amsterdam: Elsevier-Academic Press; 2015:113-132.

Staff AC, Benton SJ, von Dadelszen P, et al. Redefining preeclampsia using placenta-derived biomarkers. *Hypertension*. 2013;61(5):932. [PMID: 23460278]

Preterm Birth

Adams Waldorf KM, Singh N, Mohan AR, et al. Uterine overdistention induces preterm labor mediated by inflammation: observations in pregnant women and nonhuman primates. *Am J Obstet Gynecol*. 2015;213:830.e1. [PMID: 26284599]

Byrns MC. Regulation of progesterone signaling during pregnancy: implications for the use of progestins for the prevention of preterm birth. *J Steroid Biochem Mol Biol*. 2014;139:173. [PMID: 23410596]

Committee on Practice Bulletins—Obstetrics, The American College of Obstetricians and Gynecologists. Practice bulletin no. 130: prediction and prevention of preterm birth. *Obstet Gynecol*. 2012;120:964. [PMID: 22996126]

Menon R, Fortunato SJ, Edwards DR, Williams SM. Association of genetic variants, ethnicity and preterm birth with amniotic fluid cytokine concentrations. *Ann Hum Genet*. 2010;74:165. [PMID: 20369436]

Vrachnis N, Karavolos S, Iliodromiti Z, et al. Impact of mediators present in amniotic fluid on preterm labour. *In Vivo*. 2012;26:799. [PMID: 22949593]

This page intentionally left blank

Pancreatic Hormones and Diabetes Mellitus

Umesh Masharani, MB, BS, MRCP (UK) and
Michael S. German, MD

ABCC8	ATP-binding cassette transporter sub-family C member 8
ACE	Angiotensin-converting enzyme
ADA	American Diabetes Association
ADH	Antidiuretic hormone (vasopressin)
AGE	Advanced glycation end product
AGPAT	1-acylglycerol-3-phosphate-*O*-acyltransferase 2
AKT/PKB	AKR mouse tumor 8 kinase/protein kinase B
AIRE	Autoimmune regulator
ALT	Alanine aminotransferase
AMPK	Adenosine monophosphate-activated protein kinase
APS1	Autoimmune polyendocrinopathy syndrome type 1
ATF6	Activating transcription factor 6
BMI	Body mass index
cAMP	Cyclic adenosine monophosphate
CCK	Cholecystokinin
CEL	Carboxyl-ester lipase
cGMP	Cyclic 3′,5′-guanosine monophosphate
CIDEC	Cell death-inducing DFFA-like effector C
CSII	Continuous subcutaneous insulin infusion
CST	Contraction stress test
CTLA4	Cytotoxic T-lymphocyte-associated protein 4
DCCT	Diabetes Control and Complications Trial
DIDMOAD	Diabetes insipidus, diabetes mellitus, optic atrophy, deafness (Wolfram syndrome)
DPP	Diabetes Prevention Program
DPP-4	Dipeptidyl peptidase 4
DPT-1	Diabetes Prevention Trial-1
EIF2AK3	Eukaryotic translation initiation factor 2-α kinase 3

ER	Endoplasmic reticulum
FDA	Food and Drug Administration
FFA	Free fatty acid
FHR	Fetal heart rate
Foxo1	Forkhead box, subfamily O, member 1
FOXP3	Forkhead box, subfamily P, member 3
GABA	Gamma-aminobutyric acid
GAD	Glutamic acid decarboxylase
GCK	Glucokinase
GH	Growth hormone
GHb	Glycohemoglobin
GHSR	Growth hormone-secretagogue receptor
GI	Glycemic index
GIP	Gastric inhibitory polypeptide
GLIS3	Glioma-associated oncogene homolog—similar 3
GLP-1	Glucagon-like peptide-1
GLP-2	Glucagon-like peptide-2
GLUT	Glucose transporter
GPCR	G protein–coupled receptor
GRPP	Glicentin-related polypeptide
GWAS	Genome-wide association study
HDL	High-density lipoprotein
HLA	Human leukocyte antigen
HNF	Hepatocyte nuclear factor
hPL	Human placental lactogen
IA2	Insulinoma antigen 2
IAA	Insulin autoantibody
IAPP	Islet amyloid polypeptide
ICA	Islet cell antibody
IFG	Impaired fasting glucose
IGF-1	Insulin-like growth factor 1
IGT	Impaired glucose tolerance

IL-6	Interleukin-6
INS	Insulin
IP3	Inositol 1,4,5-triphosphate
IPEX	Immunodysregulation polyendocrinopathy enteropathy, X-linked
IPF-1	Insulin promoter factor-1
IRE1	Inositol-requiring enzyme 1
IRS	Insulin receptor substrate
ISL1	Islet transcription factor 1
IUGR	Intrauterine growth restriction
KPD	Ketosis-prone diabetes
KCNJ11	Potassium inwardly-rectifying channel, subfamily J, member 11
Kir6.2	Potassium channel, inwardly-rectifying 6.2
LADA	Latent autoimmune diabetes of adulthood
LKB1	Liver kinase B1
LDL	Low-density lipoprotein
MAPK	Mitogen-activated protein kinase
MODY	Maturity-onset diabetes of the young
MHC	Major histocompatibility complex
mTOR	Mammalian target of rapamycin
NeuroD1	Neural differentiation factor 1
Neurog3	Neural genesis factor 3
NIH	National Institutes of Health
NPH	Neutral protamine hagedorn
NST	Nonstress test
PAI-1	Plasminogen activator inhibitor-1
PAX4	Paired homeobox 4
PCSK	Proprotein convertase, subtilisin/kexin
PD-1	Programmed death 1
PD-L1	Programmed death ligand 1
PDE5	Phosphodiesterase type 5
PDX1	Pancreatic duodenal homeobox-1

PERK	Protein kinase R-like endoplasmic reticulum kinase
PGC1α	PPAR gamma coactivator-1α
PNDM	Permanent neonatal diabetes mellitus
POEMS	Polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes
PP	Pancreatic polypeptide
PPAR	Peroxisome proliferator-activated receptor
PTF1A	Pancreatic transcription factor 1α
PTP1b	Protein tyrosine phosphatase 1b
RFX6	Regulatory factor X-6
RXR	9-*cis*-retinoic acid receptor
SCL19A2	Solute carrier family 19 (thiamine transporter), member 2
SOCS	Suppressor of cytokine signaling
SREBP1c	Sterol regulatory element-binding protein 1c
SSTR	Somatostatin receptor
SUR1	Sulfonylurea receptor 1
TCF7L2	Transcription factor 7-like 2
TGF-β	Transforming growth factor-β
TNDM	Transient neonatal diabetes mellitus
TNF-α	Tumor necrosis factor-α
Ucn3	Urocortin3
UGDP	University Group Diabetes Program
UKPDS	United Kingdom Prospective Diabetes Study
VEGF	Vascular endothelial growth factor
VLDL	Very low density lipoprotein
VNTR	Variable number of tandem repeats
WFS1	Wolfram syndrome protein 1
ZAC	Zinc finger protein-inducing apoptosis and cell cycle arrest
ZnT8	Zinc transporter 8
ZFP57	Zinc finger protein 57

THE ENDOCRINE PANCREAS

The pancreas comprises two functionally distinct organs: the **exocrine pancreas**, the major digestive gland of the body; and the **endocrine pancreas**, the source of insulin, glucagon, somatostatin, pancreatic polypeptide (PP), and ghrelin. Whereas the major role of the products of the exocrine pancreas (the digestive enzymes) is the processing of ingested foodstuffs so that they become available for absorption, the hormones of the endocrine pancreas modulate every other aspect of cellular nutrition from rate of adsorption of foodstuffs to cellular storage or metabolism of nutrients. Dysfunction of the endocrine pancreas or abnormal responses to its hormones by target tissues cause serious disturbances in nutrient homeostasis, including the important clinical syndromes grouped under the name **diabetes mellitus**.

ANATOMY AND HISTOLOGY

The endocrine pancreas consists of approximately 1 million small endocrine glands—the islets of Langerhans—scattered throughout the glandular substance of the exocrine pancreas. The exocrine pancreas consists of the enzyme-producing cells organized into acini, and the duct system that delivers those enzymes to the lumen of the duodenum. The islet volume comprises 1% to 1.5% of the total mass of the pancreas and weighs about 1 to 2 g in adult humans. At least five cell types—α, β, δ, ε, and PP—have been identified in the islets (Table 17–1). Each of these islet cell types produces a distinguishing peptide hormone: glucagon, insulin, somatostatin, ghrelin, and PP, respectively. Within individual islets, the different cell types are scattered throughout. A typical human islet is depicted in Figure 17–1.

TABLE 17–1   Cell types in adult human pancreatic islets of Langerhans.

Cell Types	Approximate Percentage of Islet Volume	Secretory Products
α Cell	25	Glucagon, proglucagon
β Cell	55	Insulin, C peptide, proinsulin, IAPP, Ucn3, γ-aminobutyric acid (GABA)
δ Cell	10	Somatostatin-14
ε Cell	3	Ghrelin
PP cell	5	Pancreatic polypeptide

These cell types are not distributed uniformly throughout the pancreas. The PP cells reside primarily in islets in the posterior portion (posterior lobe) of the head, a discrete lobe of the pancreas separated from the anterior portion by a fascial partition. This lobe originates in the primordial ventral bud as opposed to the dorsal bud. The posterior lobe receives its blood supply from the superior mesenteric artery; the remainder of the pancreas derives most of its blood flow from the celiac artery. The islets themselves are richly vascularized, receiving five to ten times the blood flow
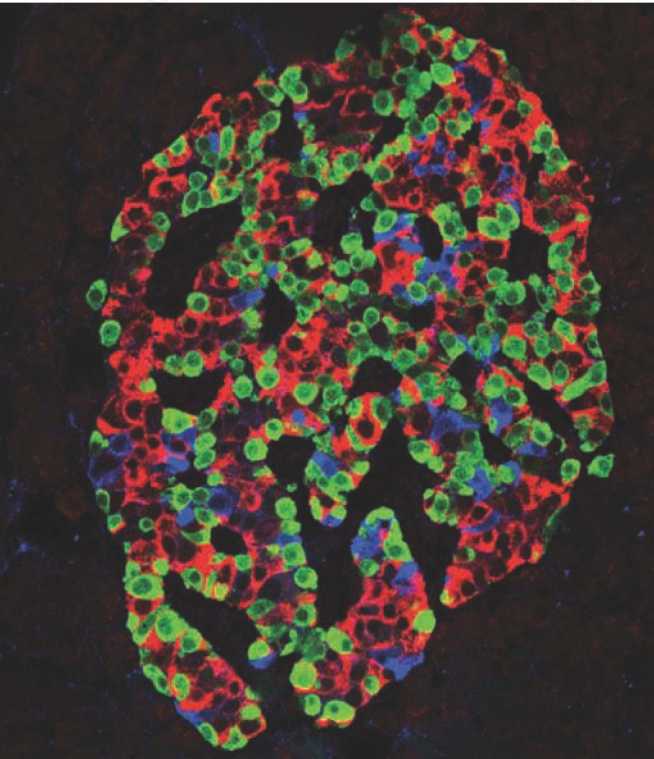
FIGURE 17–1   Human islet of Langerhans. Staining for insulin (red), glucagon (green), and somatostatin (blue) was performed by immunofluorescence and imaged by confocal microscopy. (Reproduced with permission from Cabrera O, Berman DM, Kenyon NS, et al. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci U S A.* 2006 Feb 14;103(7):2334-2339.)

of the surrounding exocrine pancreatic tissue. Each islet is surrounded by a lattice of astroglial cells and innervated by sympathetic, parasympathetic, and sensory neurons.

HORMONES OF THE ENDOCRINE PANCREAS

1. INSULIN

Biosynthesis

The human insulin gene resides on the short arm of chromosome 11. A unique set of transcription factors found in the β cell nucleus activates the transcription of the preproinsulin mRNA from the *insulin* gene (Figure 17–2). A precursor molecule, **preproinsulin**, a peptide of MW 11,500, is translated from the preproinsulin messenger RNA in the rough endoplasmic reticulum of pancreatic β cells (see Figure 17–2). Microsomal enzymes cleave preproinsulin to proinsulin (MW ~9000) almost immediately after synthesis. Proinsulin is transported to the Golgi apparatus, where packaging into clathrin-coated secretory granules takes place. Maturation of the secretory granule is associated with loss of the clathrin coating and conversion of proinsulin into insulin and a smaller connecting peptide, or C peptide, by proteolytic cleavage at two sites along the peptide chain. Mature secretory granules contain insulin and C peptide in equimolar amounts and only small quantities of proinsulin, a small portion of which consists of partially cleaved intermediates.

Biochemistry

Proinsulin (Figure 17–3) consists of a single chain of 86 amino acids, which includes the A and B chains of the insulin molecule plus a connecting segment of 35 amino acids. Two proteins—the prohormone-converting enzymes type 1 and 2 (PCSK1 and PCSK2)—are packaged with proinsulin in the immature secretory granules. These enzymes recognize and cut at pairs of basic amino acids, thereby removing the intervening sequence. After the two pairs of basic amino acids are removed by carboxypeptidase E, the result is a 51 amino acid insulin molecule and a 31 amino acid residue, the C peptide, as shown in Figure 17–3.

A small amount of proinsulin produced by the pancreas escapes cleavage and is secreted intact into the bloodstream, along with insulin and C peptide. Most anti-insulin sera used in the standard immunoassay for insulin cross-react with proinsulin; about 3% to 5% of immunoreactive insulin extracted from human pancreas is actually proinsulin. Because proinsulin is not removed by the liver, it has a half-life three to four times that of insulin. Its long half-life allows proinsulin to accumulate in the blood, where it accounts for 12% to 20% of the circulating immunoreactive insulin in the basal state in humans. Human proinsulin has about 7% to 8% of the biologic activity of insulin. The kidney is the principal site of proinsulin degradation.

Of the two major proinsulin split products present in plasma, the one split at arginine 32-33 far exceeds in amount the barely detectable 65-66 split product. In control subjects, concentrations

FIGURE 17–2 Structural components of the pancreatic β cell involved in glucose-induced biosynthesis and release of insulin. (Modified with permission from Junqueira LC, Carneiro J, Long JA. *Basic Histology*, 5th ed. New York: McGraw-Hill Education; 1986.)

FIGURE 17–3 Structure of human proinsulin C peptides and insulin molecules connected at two sites by dipeptide links.

of proinsulin and 32-33 split proinsulin after an overnight fast averaged 2.3 and 2.2 pmol/L, respectively, with corresponding postprandial rises to 10 and 20 pmol/L.

C peptide, the 31 amino acid peptide (MW 3000) released during cleavage of insulin from proinsulin, has no known biologic activity. β Cells release C peptide in equimolar amounts with insulin. It is not removed by the liver but is degraded or excreted chiefly by the kidney. It has a half-life three to four times that of insulin. In the basal state after an overnight fast, the average concentration of C peptide may remain as high as 1000 pmol/L.

Insulin is a protein consisting of 51 amino acids contained within two peptide chains: an A chain, with 21 amino acids; and a B chain, with 30 amino acids. The chains are connected by two disulfide bridges as shown in Figure 17–3. In addition, an intra-chain disulfide bridge links positions 6 and 11 in the A chain. Human insulin has a molecular weight of 5808.

Endogenous insulin has a circulatory half-life of 3 to 5 minutes. It is degraded chiefly by insulinases in liver, kidney, and placenta. A single pass through the liver removes approximately 50% of the plasma insulin.

Secretion

The human pancreas secretes about 30 units of insulin per day into the portal circulation of normal adults in distinct pulses with a period of approximately 5 minutes. The basal concentration of insulin in the peripheral blood of fasting humans averages 10 μU/mL (0.4 ng/mL, or 61 pmol/L). In normal control subjects, insulin seldom rises above 100 μU/mL (610 pmol/L) after standard meals. After ingestion of food, peripheral insulin concentration increases within 8 to 10 minutes, reaches peak concentrations by

30 to 45 minutes, and then rapidly declines to baseline values by 90 to 120 minutes postprandially.

Basal insulin secretion occurs in the absence of exogenous stimuli, in the fasting state. Plasma glucose levels below 80 to 100 mg/dL (4.4-5.6 mmol/L) do not stimulate insulin release, and most other physiologic regulators of insulin secretion only function in the presence of stimulatory levels of glucose. Stimulated insulin secretion occurs in response to exogenous stimuli. *In vivo*, ingested meals provide the major stimuli for insulin secretion. Glucose is the most potent stimulant of insulin release. The perfused pancreas releases insulin in two phases in response to glucose stimulation (Figure 17–4). When the glucose concentration increases suddenly, an initial short-lived burst of insulin release occurs (the **first phase**); if the glucose elevation persists, the insulin release gradually falls off and then begins to rise again to a steady level (the **second phase**). However, sustained levels of high glucose stimulation (~4 hours *in vitro* or >24 hours *in vivo*) result in a reversible desensitization of the β cell response to glucose but not to other stimuli.

The β cell senses glucose through its metabolism (Figure 17–5). Indeed, agents such as 2-deoxyglucose that inhibit the metabolism of glucose block the release of insulin. Glucose enters the pancreatic β cell by passive diffusion, facilitated by membrane proteins termed glucose transporters (GLUTs) (discussed later). Because the transporters function in both directions, and the β cell has an excess of GLUTs, the glucose concentration inside the β cell is in equilibrium with the extracellular glucose concentration. The low-affinity enzyme glucokinase catalyzes the subsequent, and rate-limiting, step in glucose metabolism by the pancreatic β cell, the phosphorylation of glucose to glucose-6-phosphate. Glucose catabolism in the β cell causes a rise in the intracellular ATP-ADP ratio. Acting through the sulfonylurea receptor (SUR1), the
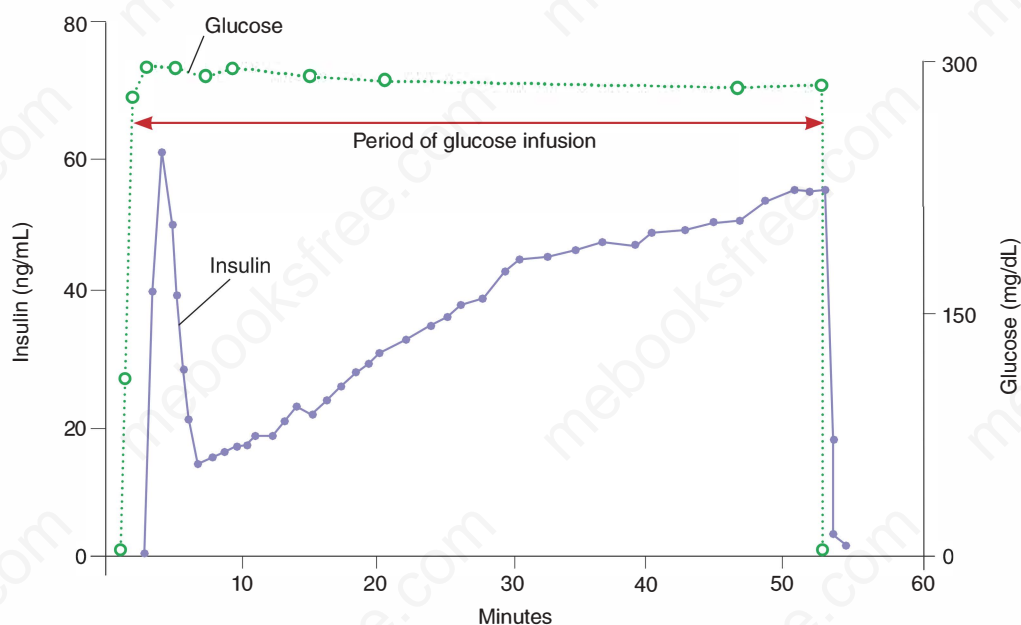
FIGURE 17–4 Multiphasic response of the *in vitro* perfused rat pancreas during constant stimulation with glucose. (Modified with permission from Grodsky GM, Curry D, Landahl H, et al. Further studies on the dynamic aspects of insulin release *in vitro* with evidence for a two-compartmental storage system. *Acta Diabetol Lat*. 1969 Sep;6 Suppl 1:554-578.)
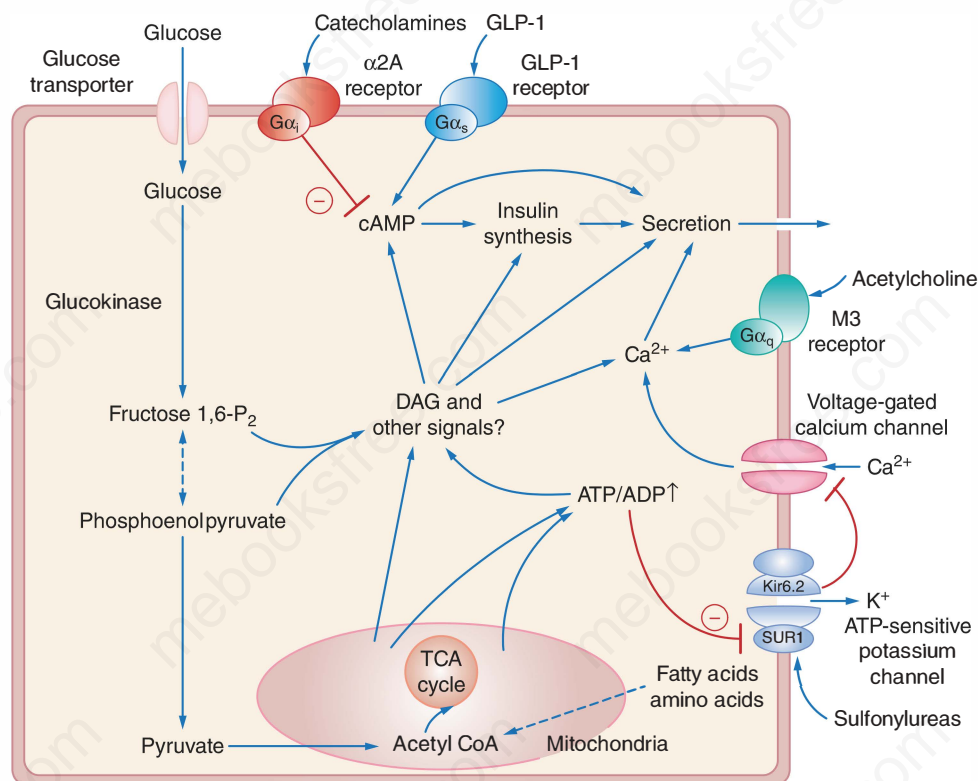
FIGURE 17–5   A simplified outline of glucose-sensing and regulated insulin secretion from the β cell. The blue arrows indicate stimulation, and the red lines indicate inhibition. Glucose enters the β cell through facultative glucose transporters, is phosphorylated to glucose-6-phosphate by glucokinse, and enters glycolysis. This results in the production of pyruvate, which enters the mitochondria, is converted to acetyl-CoA, and feeds the tricarboxylic acid (TCA) cycle and oxidative phosphorylation to produce ATP. When ATP levels rise or sulfonylureas bind to the regulatory subunit (SUR1/ABCC8) of the ATP-sensitive K^+ channels, the channel subunit (Kir6.2/KCNJ11) closes. This block of the K^+ current depolarizes the cell, allowing the voltage-gated calcium channels to open. The entry of calcium drives the fusion of insulin granules with the cell surface membrane and exocytosis of insulin. Glucose metabolism and extracellular signals modulate this pathway through release of Ca^{2+} from intracellular stores and changes in diacylglycerol (DAG), cAMP, and other intracellular signaling pathways.

nucleotide-sensing subunit of the ATP-sensitive potassium channels on the surface of the β cell, the rise in ATP-ADP ratio closes the potassium channels and depolarizes the cell, thereby activating the voltage-sensitive calcium channels and allowing the entry of calcium ions into the cell.

Insulin release requires calcium ion signaling. In addition to the voltage-dependent entry of extracellular Ca^{2+} into the β cell as described earlier, glucose also retards Ca^{2+} efflux from the β cell and releases Ca^{2+} from intracellular compartments (predominantly the endoplasmic reticulum) into the cytosol. Some nonglucose stimuli of insulin release also function through increases in cytoplasmic Ca^{2+}. The sulfonylurea and meglitinide (such as repaglinide) medications act by closing the ATP-sensitive potassium channels. Secretagogues such as acetylcholine that act through G protein–coupled receptors (GPCRs) of the $G\alpha_q$ class stimulate the release of intracellular Ca^{2+} stored in the endoplasmic reticulum by activating phospholipase C and releasing the intracellular signaling molecule inositol 1,4,5-triphosphate (IP3).

Glucose metabolism in the β cell also generates additional signals that amplify the secretory response to elevations in cytoplasmic Ca^{2+} concentration. The exact mechanisms of these amplifying signals remains unknown but involve multiple pathways and include increases in the intracellular signaling molecules diacylglycerol and cAMP. Secretagogues such as the gut hormones glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (also known as glucose-dependent insulinotropic peptide, GIP) that act via GPCRs of the $G\alpha_s$ class also stimulate insulin secretion through elevations in cAMP.

Other factors involved in the regulation of insulin secretion are summarized in Table 17–2. These factors can be divided into three categories: **direct stimulants**, which directly raise cytoplasmic calcium ion concentrations and thus can act in the absence of stimulatory glucose concentrations; **amplifiers**, which potentiate the response of the β cell to glucose; and **inhibitors**. Many of the amplifiers are **incretins**: gastrointestinal hormones that are released in response to the ingestion of meals and stimulate insulin secretion. The action of the incretins explains the observation that orally ingested glucose provokes a greater insulin secretory response than does the same amount of intravenously administered glucose.

TABLE 17–2 Regulation of insulin release.

Stimulants of Insulin Release	Glucose Amino acids: Leucine Neural: Vagal stimulation, acetylcholine Drugs: Sulfonylureas, meglitinides

Amplifiers of Glucose-Induced Insulin Release	Enteric hormones:
	  Glucagon-like peptide 1 (7-37) (GLP1)
	  Gastric inhibitory peptide (GIP)
	  Cholecystokinin, gastrin
	  Secretin
	Neural: β-adrenergic effect of
	  catecholamines
	Amino acids: arginine
	Drugs: GLP1 agonists

Inhibitors of Insulin Release	Neural: α-adrenergic effect of
	  catecholamines
	Humoral: somatostatin
	Drugs: diazoxide, thiazides, β-blockers,
	  clonidine, phenytoin, vinblastine,
	  colchicine

Insulin Receptors and Insulin Action

Insulin action begins with the binding of insulin to a receptor on the surface of the target cell membrane. Most cells of the body have specific cell surface insulin receptors. In fat, liver, and muscle cells, binding of insulin to these receptors is associated with the biologic response of these tissues to the hormone. These receptors bind insulin rapidly, with high specificity and with an affinity high enough to bind picomolar amounts.

Insulin receptors, members of the growth factor receptor family (see Chapter 1), are membrane glycoproteins composed of two protein subunits encoded by a single gene. The larger alpha subunit (MW 135,000) resides entirely extracellularly, where it binds the insulin molecule. The alpha subunit is tethered by disulfide linkage to the smaller beta subunit (MW 95,000). The beta subunit crosses the membrane, and its cytoplasmic domain contains a tyrosine kinase activity that initiates specific intracellular signaling pathways.

Downstream signaling   On binding of insulin to the alpha subunit, the beta subunit activates itself by autophosphorylation. The activated beta subunit then recruits additional proteins to the complex and phosphorylates a network of intracellular substrates, including insulin receptor substrate-1 (IRS-1), insulin receptor substrate-2 (IRS-2), and others (Figure 17–6). These activated substrates each lead to subsequent recruitment and activation of additional kinases, phosphatases, and other signaling molecules in a complex pathway that generally contains two arms: the mitogenic pathway, which mediates the growth effects of insulin and the metabolic pathway, which regulates nutrient metabolism.
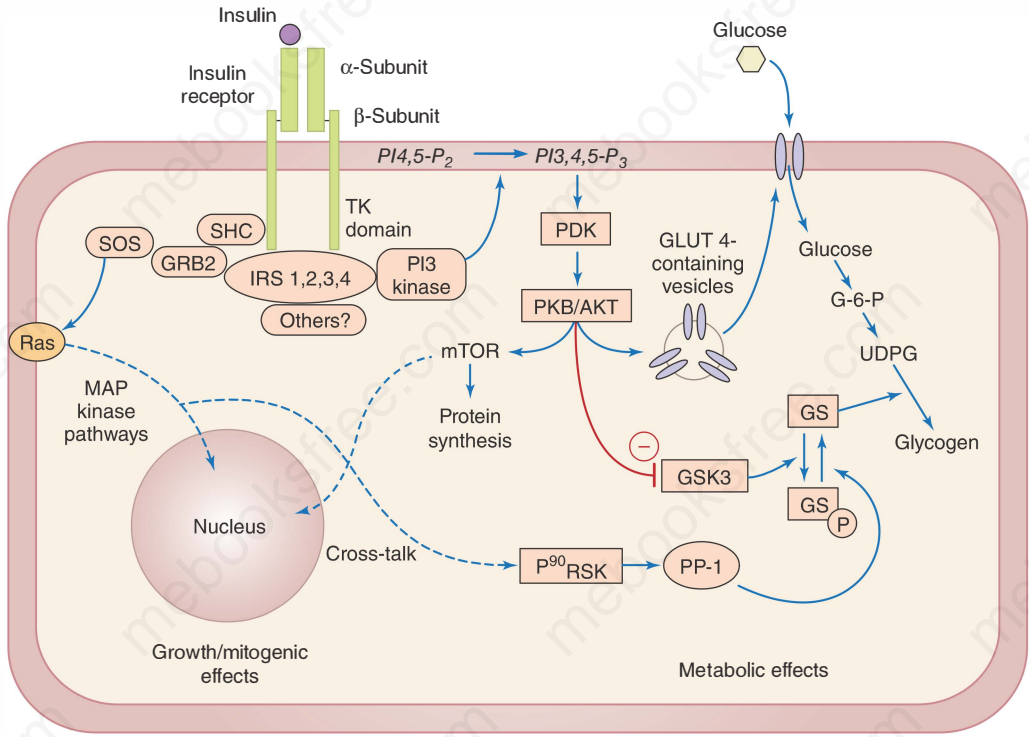
FIGURE 17–6 A simplified outline of insulin signaling. A minimal diagram of the mitogenic and metabolic arms of the insulin-signaling pathway is shown. GLUT 4, glucose transporter 4; GRB2, growth factor receptor–binding protein 2; GS, glycogen synthase (P indicates the inactive phosphorylated form); GSK3, glycogen synthase kinase 3; IRS, insulin receptor substrate (four different proteins); MAP kinase, mitogen-activated protein kinase; mTOR, mammalian target of rapamycin; PDK, phospholipid-dependent kinase; PI3 kinase, phosphatidylinositol 3 kinase; PKB/AKT, protein kinase B/AKR mouse tumor 8 kinase; PP-1, glycogen-associated protein phosphatase-1; Ras, rat sarcoma protein; SHC, Src and collagen homology protein; SOS, son-of-sevenless related protein; TK, tyrosine kinase.

In the metabolic signaling pathway, activation of phosphatidylinositol-3-kinase leads to the activation of serine/threonine kinase AKT/PKB. AKT activation drives the movement of GLUT 4–containing vesicles to the cell membrane, increases glycogen and lipid synthesis, and stimulates protein synthesis through the activation of mTOR. In the mitogenic signaling pathway, activation of Ras initiates a cascade of activating phosphorylations via the MAP kinase pathway, leading to cell growth and proliferation.

Transcriptional regulation   In addition, the insulin-signaling pathway regulates the activity of several nuclear transcription factors that in turn control the expression of genes involved in metabolism and growth. These include members of the forkhead family of transcription factors, including Foxo1, which is inactivated by phosphorylation by AKT downstream of insulin signaling. Foxo1 coordinates the expression of gene networks involved in nutrient metabolism in multiple tissues, generally activating genes involved in the response to fasting. In this process, Foxo1 works with several other transcriptional regulators including the lipogenic transcription factor SREBP1c, members of the PPAR family of nuclear receptors and the PPAR coactivator PGC1α (Figure 17–7). Foxo1 also inhibits β cell proliferation and survival.

The three members of the PPAR family of nuclear hormone receptors play pleiotropic roles in regulating genes involved in metabolism in many tissues. They may function as targets of insulin signaling, modulators of insulin signaling, or both. Despite overlap in the tissue expression and gene targets of the three PPARs, some general conclusions can be drawn about the function of each. PPARα regulates genes involved in fatty acid catabolism and gluconeogenesis and is most highly expressed in brown fat, heart, liver, kidney, and intestine. PPARβ/δ is broadly expressed and activates gene programs involved in fatty acid oxidation. PPARγ is most highly expressed in adipose tissue, intestine, and immune cells, but also at lower levels in many other tissues. PPARγ drives white adipocyte differentiation and lipid storage and inhibits production of many of the pro-resistance adipokines and pro-inflammatory cytokines in adipose tissue (see section on insulin resistance later). In macrophages, PPARγ acts to promote their alternative activation to the anti-inflammatory M2 state, rather than the pro-inflammatory M1 state.

The PPARs bind to DNA as heterodimers with the 9-*cis*-retinoic acid receptor (RXR), and recruit a variety of coactivators and corepressors. PGC1α was originally identified as a coactivator interacting with PPARγ, but the interaction is not exclusive. On different genes PPARγ works with different coactivators, and PGC1α interacts with the other PPARs and many other transcription factors. In collaboration with a variety of different transcription factors in various tissues, PGC1α orchestrates the expression of a set of genes involved in metabolism. PGC1α itself is highly regulated by several signaling pathways including insulin signaling, which inhibits PGC1α activity via phosphorylation by AKT.

A number of natural and synthetic lipids and related compounds can act as PPAR ligands, but the endogenous ligands acting *in vivo* remain a mystery. The fibrate class of lipid-lowering drugs, used clinically to lower circulating triglyceride levels, act as PPARα ligands. The thiazolidinedione class of insulin-sensitizing drugs, used for the treatment of type 2 diabetes (discussed later), act as PPARγ ligands.

Deactivation of insulin signaling   Once activated by binding to insulin, the insulin receptor and downstream signaling cascades rapidly deactivate again by several mechanisms. Insulin can simply disengage from the receptor, or the receptor can be internalized and degraded. The receptor and its tyrosine-phosphorylated substrates can be deactivated by specific protein tyrosine phosphatases such as PTP1b. In addition, inhibitory SOCS (suppressor of cytokine signaling) proteins block interactions between the phosphorylated receptor and interacting IRS proteins, direct the ubiquitination and degradation of the IRS proteins, and terminate the activation of downstream components of the signaling pathway. Finally, serine phosphorylation of the insulin receptor and its active substrates by several different serine/threonine kinases, including components of the insulin-signaling pathway such as AKT, blocks insulin signaling. Many of these mechanisms may play a role in the development of insulin resistance (discussed later).

Metabolic Effects of Insulin

The major function of insulin is to promote storage of ingested nutrients. Although insulin directly or indirectly affects the

FIGURE 17–7 Regulation and function of Foxo1. The blue arrows indicate stimulation, and the red lines indicate inhibition. Processes activated by Foxo1 are labeled in blue, while the processes inhibited by Foxo1 are shown in red. (ACC, acetyl-CoA carboxylase; ACLY, ATP-citrate lyase; AKT, AKR mouse tumor 8 kinase; FAS, fatty acid synthase; G6P, glucose-6-phosphatase; GK, glucokinase; LPK, liver pyruvate kinase; MTTP, microsomal triglyceride transfer protein; PDH, pyruvate dehydrogenase; PEPCK, phosphoenolpyruvate carboxykinase; PGC1α, peroxisome proliferator-activated receptor gamma coactivator-1; SREBP1c, sterol regulatory element-binding protein 1c).

function of almost every tissue in the body, the discussion here will be limited to a brief overview of the effects of insulin on the major tissues specialized for energy metabolism: liver, muscle, adipose tissue, and brain. In addition, the paracrine effects of insulin will be discussed briefly.

A. Paracrine effects   The effects of the products of endocrine cells on surrounding cells are termed *paracrine effects*, in contrast to actions that take place at sites distant from the secreting cells, which are termed *endocrine effects* (see Chapter 1). Paracrine effects of the β and δ cells on the nearby α cells (see Figure 17–1) are of considerable importance in the endocrine pancreas. Insulin directly inhibits α cell secretion of glucagon. In addition, somatostatin, which δ cells release in response to most of the same stimuli that provoke insulin release, also inhibits glucagon secretion.

Because glucose stimulates only β and δ cells (whose products then inhibit α cells) whereas amino acids stimulate glucagon as well as insulin, the type and amounts of islet hormones released during a meal depend on the ratio of ingested carbohydrate to protein. The higher the carbohydrate content of a meal, the lower the amount of glucagon released by any amino acids absorbed. In contrast, a predominantly protein meal results in relatively greater glucagon secretion, because amino acids are less effective at stimulating insulin release in the absence of concurrent hyperglycemia but are potent stimulators of α cells.

B. Endocrine effects (Table 17–3)

1. **Liver**—The first major organ reached by insulin via the bloodstream is the liver. Insulin exerts its action on the liver in two major ways:
   a. **Insulin promotes anabolism**—Insulin promotes glycogen synthesis and storage while inhibiting glycogen breakdown. These effects are mediated by changes in the activity of enzymes in the glycogen synthesis pathway (discussed later). The liver has a maximum storage capacity of 100 to 110 g of glycogen, or approximately 440 kcal of energy.

   Insulin increases both protein and triglyceride synthesis and very low density lipoprotein (VLDL) formation by the liver. It also inhibits gluconeogenesis and promotes glycolysis through its effects on the function and expression of key enzymes of both pathways.
   b. **Insulin inhibits catabolism**—Insulin acts to reverse the catabolic events of the postabsorptive state by inhibiting hepatic glycogenolysis, ketogenesis, and gluconeogenesis.
2. **Muscle**—Insulin promotes protein synthesis in muscle by increasing amino acid transport, as well as by stimulating ribosomal protein synthesis. In addition, insulin promotes glycogen synthesis to replace glycogen stores expended by muscle activity. This is accomplished by increasing glucose transport into the muscle cell, enhancing the activity of glycogen synthase, and inhibiting the activity of glycogen phosphorylase. Approximately 500 to 600 g of glycogen are stored in the muscle tissue of a 70-kg man, but because of the lack of glucose 6-phosphatase in this tissue, it cannot be used as a source of blood glucose, except for a small amount produced when the debranching enzyme releases unphosphorylated glucose from branch points in the glycogen polymer, and the

TABLE 17–3 Endocrine effects of insulin.

Tissue	Effect of Insulin
Liver	**Catabolic Pathways** Inhibits glycogen breakdown (inhibits glycogen phosphorylase) Inhibits conversion of fatty acids and amino acids to keto acids Inhibits conversion of amino acids to glucose (gluconeogenesis) **Anabolic Pathways** Promotes glucose storage as glycogen (induces glucokinase and glycogen synthase) Increases triglyceride synthesis and VLDL formation
Muscle	**Protein Synthesis** Increases amino acid transport Increases ribosomal protein synthesis **Glycogen Synthesis** Increases glucose transport Induces glycogen synthase Inhibits glycogen phosphorylase
Adipose Tissue	**Triglyceride Storage** Lipoprotein lipase is induced by insulin to hydrolyze triglycerides in circulating lipoproteins for delivery of fatty acids to the adipocytes Glucose transport into cell provides glycerol phosphate to permit esterification of fatty acids supplied by lipoprotein transport Intracellular lipase is inhibited by insulin
Brain	**Decreased appetite** **Increased energy expenditure**

glucose indirectly produced via the liver from lactate generated by muscle.
3. **Adipose tissue**—Fat, in the form of triglyceride, is the most efficient means of storing energy. It provides 9 kcal/g of stored substrate, as opposed to the 4 kcal/g generally provided by protein or carbohydrate. In the typical 70-kg man, the energy content of adipose tissue is about 100,000 kcal.

   Insulin acts to promote triglyceride storage in adipocytes by a number of mechanisms. (1) It induces the production of lipoprotein lipase in adipose tissue (this is the lipoprotein lipase that is bound to endothelial cells in adipose tissue and other vascular beds), which leads to hydrolysis of triglycerides from circulating lipoproteins, thereby yielding fatty acids for uptake by adipocytes. (2) By increasing glucose transport into fat cells, insulin increases the availability of α-glycerol phosphate, a substance used in the esterification of free fatty acids into triglycerides. (3) Insulin inhibits intracellular lipolysis of stored triglyceride by inhibiting intracellular lipase (also called **hormone-sensitive lipase**). This reduction of fatty acid flux to the liver is a key regulatory factor in the action of insulin to lower hepatic gluconeogenesis and ketogenesis.
4. **Central nervous system**—Although the brain is traditionally not considered an insulin-sensitive tissue, and overall glucose utilization by the brain is not acutely regulated by insulin, key regions of the brain can respond to insulin. Insulin signaling via PI3 kinase in key cells in the hypothalamus functions with leptin signaling to decrease appetite and increase energy expenditure (see Chapter 20).

C. AMPK and insulin-independent regulation of nutrient metabolism

Insulin, along with the counter-regulatory hormones and other circulating enhancers and inhibitors of their actions, coordinates nutrient metabolism in response to the overall needs of the organism. At the level of the individual cell, however, additional mechanisms sense and respond to the local energy state. Among these mechanisms, adenosine monophosphate protein kinase (AMPK) plays a central role. When energy availability falls, the drop in cellular ATP concentration and rise in AMP trigger a conformational change in the trimeric AMPK complex and the subsequent activation of the catalytic domain by the serine/threonine kinase LBK1/STK11. AMPK then drives the production of ATP by activating catabolic pathways and inhibiting synthetic pathways in the cell (Figure 17–8). In muscle, in response to the rise in AMP during exercise, AMPK increases fatty acid oxidation and insulin-independent glucose uptake while inhibiting mTOR and protein synthesis. In the long term, AMPK also drives mitochondrial biogenesis. In liver cells, AMPK blocks fatty acid and triglyceride synthesis while activating fatty acid oxidation, and also inhibits the gluconeogenic program by blocking cAMP activation of gene expression and inhibiting Foxo1/PGC1α-driven expression of the gluconeogenic genes. In brain, AMPK also functions as an energy sensor and plays a role in the regulation of appetite and energy expenditure by the hypothalamus. AMPK has also been implicated in the regulation of insulin secretion by β cells.

While predominantly an intracellular energy sensor, AMPK increases the sensitivity of cells to insulin, although the mechanisms remain uncertain. AMPK also responds to extracellular signals, and contributes to the regulation of metabolism by many of the adipokines and cytokines (discussed later) as well as cannabinoids. The biguanide drugs, including metformin, which is used in the treatment of type 2 diabetes, activate AMPK by reducing mitochondrial production of ATP and raising intracellular levels of AMP, and thereby lower blood glucose levels by inhibiting gluconeogenesis.

Glucose Transporter Proteins

Glucose oxidation provides energy for most cells and is critical for brain function. Because cell membranes are impermeable to hydrophilic molecules such as glucose, all cells require carrier proteins to transport glucose across the lipid bilayers into the cytosol. All cells utilize non-energy-dependent transporters that facilitate diffusion of glucose from a higher concentration to a lower concentration across cell membranes. Facilitative GLUTs comprise a large family including at least 13 members, although some of the recently identified members of the family have not yet been shown to transport glucose. The first four members of the family are the best characterized, and they have distinct affinities for glucose and distinct patterns of expression.
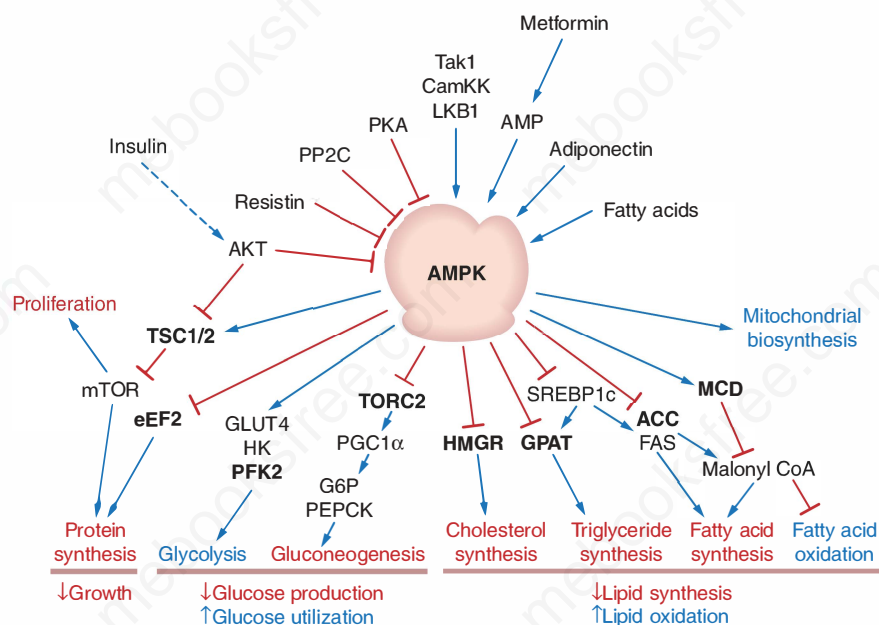
FIGURE 17–8 Regulation and function of AMPK. Proteins that are directly phosphorylated by AMPK are shown in bold font. The blue arrows indicate stimulation, and the red lines indicate inhibition. Processes activated by AMPK are labeled in blue, while the processes inhibited by AMPK are shown in red. (ACC, acetyl-CoA carboxylase; AKT, AKR mouse tumor 8 kinase; CamKK, calcium/calmodulin-dependent protein kinase kinase; eEF2, eukaryotic translation elongation factor 2; FAS, fatty acid synthase; G6P, glucose-6-phosphatase; GPAT, glycerol-3-phosphate acyltransferase, mitochondrial; HK, hexokinase; HMGR, HMG-CoA reductase; LKB1, liver kinase B1; MCD, malonyl-CoA decarboxylase; mTOR, mammalian target of rapamycin; PEPCK, phosphoenolpyruvate carboxykinase; PFK2, 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase; PGC1α, peroxisome proliferator-activated receptor gamma coactivator-1; PKA, protein kinase A; PP2C, protein phosphatase 2C; SREBP1c, sterol regulatory element–binding protein 1c; Tak1, TGF-β-activated kinase 1; TORC2, transducer of regulated cAMP response element-binding protein 2; TSC1/2, tuberous sclerosis 1).

GLUT 1 is present in all human tissues. It mediates basal glucose uptake, because it has a very high affinity for glucose and, therefore, can transport glucose at the relatively low concentrations found in the fasted state. For this reason, its presence on the surface of the endothelial cells of the brain vascular system (blood–brain barrier) ensures adequate transport of plasma glucose into the central nervous system.

GLUT 3, which is also found in all tissues, is the major GLUT on neurons. It also has a very high affinity for glucose and is responsible for transferring glucose into neuronal cells at the lower concentrations found in the central nervous system.

In contrast, GLUT 2 has a lower affinity for glucose and thus increases glucose transport when plasma glucose levels rise, such as postprandially. It is a major transporter of glucose in hepatic, intestinal, and renal tubular cells. The low affinity of GLUT 2 for glucose reduces hepatic uptake of glucose during fasting, while its ability to transport glucose equally efficiently in both directions assists in the export of glucose from hepatocytes. GLUT 2 is also expressed on the surface of the β cells in rodents, but it is not detected at significant levels on human β cells.

GLUT 4 is found in two major insulin target tissues: skeletal muscle and adipose tissue. It is sequestered mainly within an intracellular compartment of these cells and thus does not function as a GLUT until insulin signaling causes translocation of GLUT 4 to the cell membrane, where it facilitates glucose entry into these tissues after a meal (see Figure 17–6). In muscle, exercise also drives GLUT 4 translocation to the cell surface by activating AMPK.

The intestine and kidney also have energy-dependent Na^+-glucose cotransporters capable of transporting glucose against its concentration gradient. SGLT1 transports glucose from the gut lumen into the absorptive enterocytes lining the lumen and is also found in the proximal tubules of the kidney. SGLT2, however, which is found in the proximal convoluted tubules in the kidney, performs the majority of the renal glucose reabsorption.

Islet Amyloid Polypeptide

Islet amyloid polypeptide (IAPP), or amylin, is a 37 amino acid peptide produced and stored with insulin in pancreatic β cells but only at a low ratio of approximately one molecule of IAPP to 100 of insulin. β cells cosecrete IAPP with insulin in response to glucose and other β cell secretagogues. Although it plays a role in regulating gut physiology by decreasing gastric emptying and gut motility after meals, the full physiologic functions of IAPP remain uncertain. A soluble analog of IAPP called pramlintide has been approved for use in patients with type 1 diabetes and insulin-treated type 2 diabetes (discussed later).

IAPP produces amyloid deposits in pancreatic islets of most patients with type 2 diabetes of long duration. These amyloid deposits are insoluble fibrillar proteins generated from IAPP oligomers that encroach on and may even occur within pancreatic β cells. Islets of nondiabetic elderly persons may contain less extensive amyloid deposits. Whether amyloid fibrils and deposition contribute to the islet dysfunction and β cell loss seen in type 2 diabetes or are simply a consequence of disordered and hyperstimulated islet function remains an unresolved question.

2. GLUCAGON
Biochemistry

Pancreatic glucagon, along with several other biologically active peptides, derives from the large proglucagon peptide encoded by the preproglucagon gene located on human chromosome 2. Tissue-specific proteases (the prohormone convertases) cleave different sets of peptide products from the proglucagon molecule in the endocrine L-cells of the gut and the α cells in the islet (Figure 17–9). The activity of prohormone convertase 2 in α cells generates the glucagon peptide, along with the amino-terminal glicentin-related peptide, a small central hexapeptide, and a large carboxyl-terminal fragment.

Glucagon consists of 29 amino acids in a single-chain polypeptide with a molecular weight of 3485. In healthy humans, the average fasting plasma immunoreactive glucagon level is 75 pg/mL (25 pmol/L). Only 30% to 40% of this is actually pancreatic glucagon, the remainder being a heterogeneous composite of higher-molecular-weight molecules with glucagon immunoreactivity such as proglucagon, glicentin, and oxyntomodulin. Circulating glucagon has a half-life of 3 to 6 minutes due to removal by the liver and kidney.

Secretion

In contrast to its stimulation of insulin secretion, glucose inhibits glucagon secretion. Conflicting data surround the question of whether glucose directly inhibits secretion from the α cell or whether it only acts via release of insulin and somatostatin from the β and δ cells, both of which inhibit the α cell directly. In addition, because β cells release gamma-aminobutyric acid (GABA) and α cells express inhibitory GABA receptors, GABA also may participate in the inhibition of α cells during β cell stimulation.

Many amino acids stimulate glucagon release, although they differ in their ability to do so. Some, such as arginine, release both glucagon and insulin; others (eg, alanine) stimulate primarily glucagon release. Leucine, an effective stimulant of insulin release, does not stimulate glucagon. Other substances that promote glucagon release include catecholamines, gastrointestinal hormones (cholecystokinin [CCK], gastrin, and gastric inhibitory polypeptide [GIP]), and glucocorticoids. Both sympathetic and parasympathetic (vagal) stimulation promote glucagon release, especially in response to hypoglycemia. High levels of circulating fatty acids suppress glucagon secretion.

Action of Glucagon

In contrast to insulin, which promotes energy storage in a variety of tissues in response to feeding, glucagon provides a humoral mechanism for delivering energy from the liver to the other tissues between meals. The ratio of insulin to glucagon affects key target tissues by regulating the expression and activity of key enzymes controlling nutrient metabolism and, thereby, controlling the flux of these nutrients into or out of storage.

The liver, because of its connection to the pancreas via the portal vein, represents the major target organ for glucagon, with
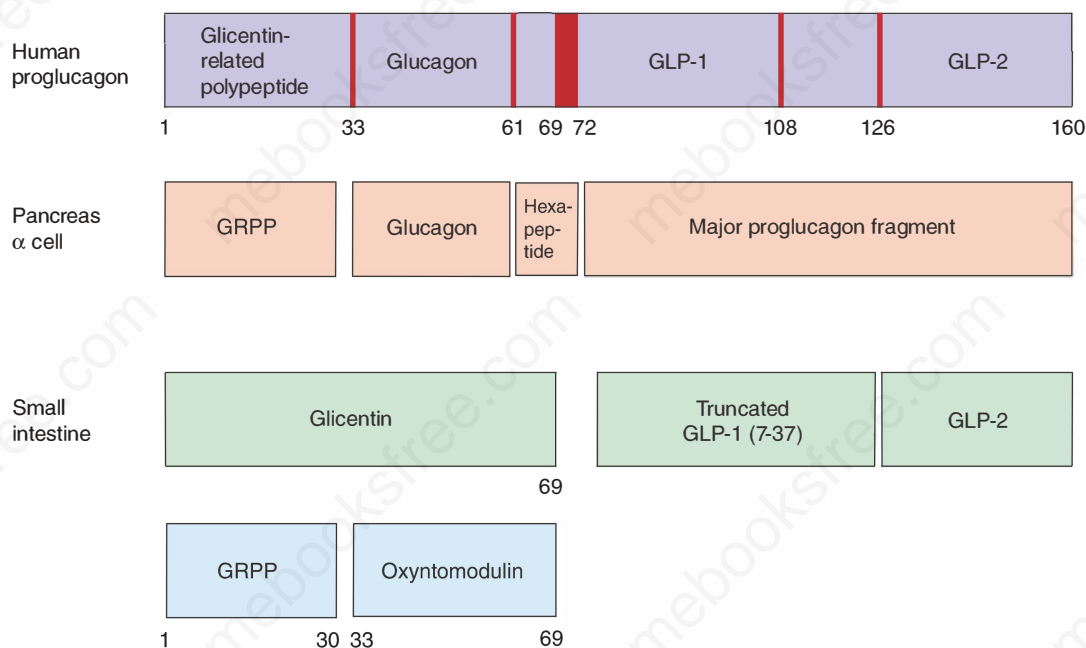
FIGURE 17–9 Tissue-specific secretory products of human proglucagon (GLP-1, glucagon-like peptide-1; GLP-2, glucagon-like peptide-2; GRPP, glicentin-related polypeptide).

portal vein glucagon concentrations reaching as high as 300 to 500 pg/mL (100-166 pmol/L) during fasting. It is unclear whether physiologic levels of glucagon affect tissues other than the liver. Glucagon signals through the glucagon receptor, a GPCR of the $G\alpha_s$ class found predominantly on the surface of hepatocytes. Binding of glucagon to its receptor in the liver activates adenylyl cyclase and the generation of cAMP, which in turn mediates the phosphorylation or dephosphorylation of key enzymes regulating nutrient metabolism. In addition, like insulin, glucagon receptor signaling modifies the activity of a set of cAMP responsive transcriptional regulators that in turn control the expression of the genes encoding these same enzymes.

Glucagon signaling in the liver stimulates the breakdown of stored glycogen, maintains hepatic output of glucose from amino acid precursors (gluconeogenesis), and promotes hepatic output of ketones generated from fatty acid precursors (ketogenesis). Glucagon facilitates the uptake of the gluconeogenic substrate alanine by liver, and directs fatty acids away from reesterification to triglycerides and toward ketogenic pathways. In sum, glucagon signaling results in the net release of readily available energy stores from the liver in the form of glucose and ketones.

Glucagon-Related Peptides

In the intestinal ʟ-cells, found predominantly in the distal ileum and colon, prohormone convertase 1 generates a different set of peptides from the proglucagon molecule, including glicentin, glicentin-related polypeptide (GRPP), oxyntomodulin, and the two glucagon-like peptides **GLP-1** and **GLP-2** (see Figure 17–9). Several biological activities have been attributed to glicentin and oxyntomodulin based on studies using high concentrations of the peptides, but all these actions can be explained by low-affinity interactions with the receptors for glucagon, GLP-1 and GLP-2. Specific receptors for glicentin and oxyntomodulin have not been identified, and it remains uncertain whether these peptides play any biological role at physiologic concentrations. GRPP also has no clearly established biological activity. The other two gut-derived glucagon-related peptides, GLP-1 and GLP-2, however, play important roles in nutrient metabolism and gastrointestinal physiology (Table 17–4).

There are two active forms of GLP-1: GLP-1(7-36) amide, and GLP-1(7-37). The intestinal ʟ-cells secrete GLP-1 in response to meals, through dietary glucose and lipids and parasympathetic stimulation. The ʟ-cells sense dietary fat in the gut lumen in part through the GPR119 receptor, which binds the long chain fatty acid derivative oleoylethanolamide. GPR119 is also expressed on the surface of the β cells. GLP-1 binds to the GLP-1 receptor, a GPCR similar to the glucagon receptor. The ubiquitous protease dipeptidyl peptidase 4 (**DPP-4**) rapidly inactivates circulating GLP-1 (half-life <2 min) by removing the two amino-terminal amino acids. Pancreatic islets are major targets of GLP-1 action. GLP-1 directly stimulates the production and secretion of insulin and somatostatin, and thereby indirectly inhibits the secretion of glucagon. In addition, GLP-1 protects the β cells from destruction and stimulates β cell growth. Other targets of GLP-1 include the stomach, where the peptide inhibits gastric emptying and gastric acid secretion; the brain, where it inhibits appetite and induces weight loss; and the heart, where it has some protective effects.

Along with GLP-1, intestinal ʟ-cells cosecrete GLP-2 in response to eating; and like GLP-1, GLP-2 binds to a specific GPCR closely related to the glucagon and GLP-1 receptors. DPP-4 also inactivates GLP-2. GLP-2 signaling predominantly targets the intestine, where it stimulates mucosal growth and nutrient absorption and inhibits motility.

TABLE 17–4 **Biologic roles of glucagon-related peptides.**

Target Tissue	Glucagon	GLP-1	GLP-2	GIP
Islet	**Stimulates** insulin secretion	**Stimulates** insulin and soma-tostatin secretion **Inhibits** glucagon secretion (indirectly) **Inhibits** β cell death		**Stimulates** insulin, somatosta-tin, and glucagon secretion **Inhibits** glucagon secretion (indirectly) **Inhibits** β cell death
Liver	**Stimulates** glycogenolysis, gluconeogenesis, fatty acid oxidation, and ketogenesis **Inhibits** glycogen synthesis and fatty acid synthesis			
Stomach		**Inhibits** gastric acid secretion **Inhibits** gastric emptying		**Inhibits** gastric acid secretion and gastric emptying
Intestine			**Stimulates** mucosal growth and nutri-ent absorption **Inhibits** motility	**Stimulates** GLP-1 secretion by L-cells
Adipose tissue				**Stimulates** adipogenesis, lipogenesis, and adipokine production
Brain (hypothalamus)		**Inhibits** appetite		

The K cells in the duodenum and jejunum produce a related 42 amino acid incretin peptide, **GIP**, that has both functional and sequential similarity to GLP-1, but is the product of a distinct gene and binds to a distinct receptor, GIPR, which also belongs to the family of glucagon-related $G\alpha_s$-linked receptors. The K cells secrete GIP in response to glucose—via the same pathway used by the β cell (see Figure 17–5)—and lipids. Interestingly, the GIP prepropeptide is also expressed in α cells, but prohormone conver-tase 2 in α cells produces a shorter peptide, GIP_{1-30}, which lacks the 12 carboxyl amino acids present in intestinal GIP_{1-42}. The two forms of GIP appear to function identically. GIP signaling through its receptor has similar effects to those of GLP-1 on the stomach and β cells. α Cells also express the GIP receptor, through which GIP directly stimulates glucagon secretion; but GIP concomitantly suppresses glucagon secretion indirectly

through its stimulation of insulin secretion. The GIP receptor is also expressed in adipose tissue and bone. In adipose tissue, GIP plays an important role in the differentiation of new adipocytes, and also drives lipogenesis and adipokine production in mature adipocytes. In bone, GIP stimulates the osteoblasts and increases bone density.

3. SOMATOSTATIN

The pancreatic δ cells transcribe the gene for somatostatin on the long arm of chromosome 3. It codes for a 116 amino acid peptide, preprosomatostatin, from whose carboxyl end is cleaved the hor-mone somatostatin, a 14 amino acid cyclic polypeptide with a molecular weight of 1640 (Figure 17–10). First identified in the hypothalamus, it owes its name to its ability to inhibit the release
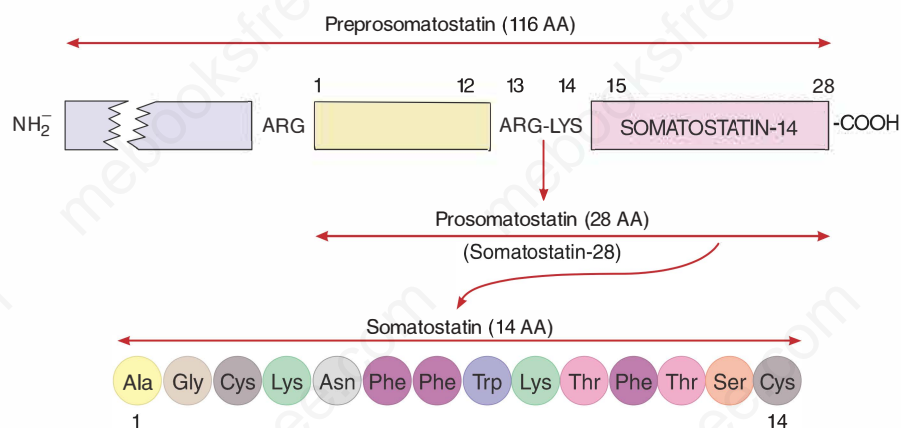
FIGURE 17–10 Amino acid sequence of somatostatin and its cleavage from dibasic amino acid residue in prosomatostatin and preprosomatostatin.

of growth hormone (GH; pituitary somatotropin). Since that time, somatostatin has been found in a number of tissues, including many areas of the brain and peripheral nervous system, the endocrine D cells in the epithelial lining of the stomach and intestine, and the δ cells in the pancreatic islets. In neurons, gastric D cells and the islet, somatostatin-14 predominates, but approximately 5% to 10% of the somatostatin-like immunoreactivity in the brain consists of a 28 amino acid peptide, somatostatin-28. Somatostatin-28 consists of an amino terminal region of 14 amino acids and a carboxyl terminal segment containing somatostatin-14. In small intestine, the larger molecule predominates, with 70% to 75% of the hormone in the 28 amino acid form and only 25% to 30% as somatostatin-14. Somatostatin-28 is 10 times more potent than somatostatin-14 in inhibiting growth hormone and insulin secretion, whereas somatostatin-14 is more effective in inhibiting glucagon release.

Most known stimulators of insulin release also promote somatostatin release from δ cells. These include glucose, arginine, gastrointestinal hormones, and sulfonylureas. The peptide hormone Urocortin3 (Ucn3), which is secreted by the β cells, also stimulates somatostatin secretion. The importance of circulating somatostatin is unclear; the major action of this peptide appears to be paracrine regulation of the pancreatic islet and the gastrointestinal tract. Physiologic levels of somatostatin in humans seldom exceed 80 pg/mL (49 pmol/L). The metabolic clearance of exogenously infused somatostatin in humans is extremely rapid; the half-life of the hormone is less than 3 minutes.

Molecular cloning has identified five somatostatin receptors (SSTR1-5), all of which are GPCRs. They vary in size from 364 to 418 amino acids (with 105 amino acids invariant) and function in the central nervous system and a wide variety of peripheral tissues, including the pituitary gland, the small intestine, and the pancreas. All five receptors belong to the $G\alpha_i$ class and inhibit the activity of adenylate cyclase, thereby lowering intracellular levels of cAMP and inhibiting cAMP-activated secretion. In addition, however, each of the different somatostatin receptors interacts with additional distinct downstream effectors that modify the cellular consequences of receptor activation. Binding of ligand to SSTR5 on β cells mediates the inhibition of insulin secretion, whereas inhibition of GH release from pituitary somatotrophs as well as glucagon release from α cells of the pancreas works through SSTR2. This explains why an analog of somatostatin, octreotide, which has a much greater affinity for SSTR2 than for SSTR5, is effective in correcting GH excess without having much effect on carbohydrate tolerance when used to treat acromegaly.

Somatostatin acts in several ways to restrain the movement of nutrients from the intestinal tract into the circulation. It prolongs gastric emptying time, decreases gastric acid and gastrin production, diminishes pancreatic exocrine secretion, decreases splanchnic blood flow, and retards xylose absorption.

4. PANCREATIC POLYPEPTIDE

PP is found in PP cells located chiefly in islets in the posterior portion of the head of the pancreas. Similar to the other islet hormones, PP derives from a larger prepropeptide of 85 amino acids that is cleaved to a single 36 amino acid peptide with a molecular weight of 4200. Circulating levels of the peptide increase in response to a mixed meal; however, intravenous infusion of glucose or lipid does not produce such a rise, and intravenous amino acids induce only a small increase. In contrast, vagotomy abolishes the response to an ingested meal, demonstrating that PP secretion responds predominantly to neural, rather than nutrient signals.

In healthy subjects, basal levels of PP average 24 ± 4 pmol/L and may become elevated owing to a variety of factors including old age, alcohol abuse, diarrhea, chronic renal failure, hypoglycemia, or inflammatory disorders. Values above 300 pmol/L are found in most patients with pancreatic endocrine tumors such as glucagonoma or vasoactive intestinal polypeptide-secreting tumor and in all patients with tumors of the pancreatic PP cell. As many as 20% of patients with insulinoma and one-third of those with gastrinomas also have plasma concentrations of PP that are greater than 300 pmol/L.

Although it has been implicated in the regulation of exocrine pancreatic secretion and gall bladder contraction, the physiologic actions of PP remain uncertain.

5. GHRELIN

The peptide hormone ghrelin was originally identified in extracts from the stomach based on its ability to bind to and activate the growth hormone secretagogue receptor (GHSR) and stimulate growth hormone release from the pituitary (thus the basis of its name: **g**rowth **h**ormone-**rel**easing peptide). The P/D1 endocrine cells in the gastric mucosa and the ε cells in the islet make ghrelin, as do a few cells in the heart, lung, kidney, immune system, hypothalamus, and pituitary. The human *GHRELIN* gene comprises four exons, and the major splice product encodes the 117 amino acid preproghrelin peptide. Processing in the ε cells yields the active form of ghrelin: a 28 amino acid peptide (amino acids 24-51 of preproghrelin) with the serine in position 3 modified by the attachment of an octanoyl side chain by Ghrelin O-acyltransferase (GOAT). Full biological activity requires the *n*-octanoyl modification. In addition, protease cleavage generates a second peptide, obestatin (amino acids 76-98 of preproghrelin) of less certain biological function.

Initially identified as a stimulator of growth hormone secretion, ghrelin signals through its receptor, the previously identified GHSR, which is a GPCR found in a variety of tissues, including the hypothalamus, pituitary, intestine, and islet. Ghrelin signaling stimulates growth hormone secretion directly through its receptor on pituitary somatotrophs, and also through its stimulation of hypothalamic GHRH secretion. In addition, ghrelin induces gastric emptying and acid secretion and regulates appetite and energy balance via neurons in the arcuate nucleus of the hypothalamus (see Chapter 20). Recent evidence from rodents suggests that ghrelin plays an essential role in conserving energy and maintaining blood glucose levels during starvation. The role of ghrelin signaling in the pancreas, and the relative contribution of islet-derived ghrelin to the overall actions of ghrelin remains unresolved.

DIABETES MELLITUS

Clinical diabetes mellitus is a syndrome of disordered metabolism with inappropriate hyperglycemia due to an absolute or relative deficiency of insulin. There may also be a defect in insulin action (insulin resistance).

CLASSIFICATION

Diabetes is classified into five main groups based on known pathological and etiologic mechanisms—type 1, type 2, monogenic, secondary, and gestational diabetes (Table 17–5). Type 1 diabetes (previously referred to as juvenile-onset or insulin dependent diabetes mellitus [IDDM]) results from pancreatic islet β cell destruction most commonly by an autoimmune process. These patients are prone to developing ketoacidosis and require insulin replacement. Type 2 diabetes (previously referred to as adult-onset or non-insulin-dependent diabetes mellitus [NIDDM]), the most prevalent form of diabetes, is a heterogeneous disorder most commonly associated with insulin resistance combined with an impairment in compensatory insulin secretion.

TYPE 1 DIABETES MELLITUS

Type 1 diabetes is immune-mediated in more than 95% of cases (type 1a) and idiopathic in less than 5% (type 1b). The rate of pancreatic β cell destruction may vary, but in most cases the process is prolonged, extending over months or years, since evidence for an immune response can be detected long in advance of hyperglycemia in patients that eventually develop type 1 diabetes. It is a catabolic disorder in which circulating insulin is virtually absent, plasma glucagon is elevated, and the pancreatic β cells fail to respond to all known insulinogenic stimuli. In the absence of insulin, the three main target tissues of insulin (liver, muscle, and fat)

TABLE 17–5 Etiologic classification of diabetes mellitus.

I. Type 1 Diabetes[a] (β cell destruction, usually leading to absolute insulin deficiency)
   A. Immune-mediated, type 1a
   B. Idiopathic, type 1b

II. Type 2 Diabetes[a] (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly secretory defect with minimal insulin resistance)

III. Monogenic Diabetes
   A. Autosomal dominant genetic defects of pancreatic β cells
       1. Maturity onset diabetes of the young (MODY)
       2. Insulin gene (*INS*)
       3. ATP-sensitive potassium channel (*KCNJ11* and *ABCC8*)
   B. Other genetic defects of pancreatic β cells
       1. Autosomal recessive genetic defects
       2. Mitochondrial DNA
       3. Ketosis-prone diabetes (KPD)
   C. Genetic defects in insulin action
       1. Insulin receptor mutations
       2. Lipoatrophic diabetes
   D. Neonatal diabetes (overlaps with categories A, B, C, E, and F)
       1. Transient
       2. Permanent
   E. Monogenic autoimmune syndromes
       1. IPEX: Immunodysregulation polyendocrinopathy enteropathy, X-linked
       2. Autoimmune polyendocrinopathy syndrome type 1
       3. Other autosomal recessive autoimmune polyendocrinopathies (*IL2RA*, *ITCH*, and *LRBA*)
       4. Other autosomal dominant autoimmune polyendocrinopathies (*STAT1*, *STAT3*, and *SIRT1*)
   F. Other genetic syndromes sometimes associated with diabetes
       1. Chromosomal defects: Down, Klinefelter, and Turner syndromes
       2. Neuromuscular syndromes: Friedreich ataxia, Huntington chorea, myotonic dystrophy, porphyria, and others
       3. Obesity syndromes: Laurence-Moon-Biedl, Bardet-Biedl, Prader-Willi syndromes, and others
       4. Wolfram syndrome

IV. Secondary Diabetes
   A. Diseases of the exocrine pancreas
       1. Pancreatitis
       2. Trauma, pancreatectomy
       3. Neoplasia
       4. Cystic fibrosis
       5. Hemochromatosis
       6. Fibrocalculous pancreatopathy
   B. Endocrinopathies
       1. Acromegaly
       2. Cushing syndrome
       3. Glucagonoma
       4. Pheochromocytoma
       5. Hyperthyroidism
       6. Somatostatinoma
       7. Aldosteronoma
   C. Drug- or chemical-induced
       1. β cell toxicity: vacor, pentamidine, cyclosporine
       2. β cell autoimmunity: α-interferon, anti-PD-1, anti-PD-L1, anti-CTLA-4
       3. β cell dysfunction: thiazide and loop diuretics, diazoxide, α agonists, β blockers, phenytoin, opiates
       4. Insulin resistance: glucocorticoids, progesterone, nicotinic acid, thyroid hormone, β blockers, atypical antipsychotic drugs, antiretroviral protease inhibitors
   D. Infections
       1. Congenital rubella
       2. Other viruses: cytomegalovirus, coxsackievirus B, adenovirus, mumps
   E. Uncommon forms of immune-mediated diabetes
       1. Stiff-person syndrome
       2. Anti-insulin receptor antibodies
       3. POEMS syndrome

V. Gestational Diabetes Mellitus (GDM)

[a]Patients with any form of diabetes may require insulin treatment at some stage of their disease. Such use of insulin does not, of itself, classify the patient.

Data from American Diabetes Association. Classification and Diagnosis of Diabetes. *Diabetes Care*. 2016 Jan;39 Suppl 1:S13-22.

not only fail to appropriately take up absorbed nutrients but continue to deliver glucose, amino acids, and fatty acids into the bloodstream from their respective storage depots. Furthermore, alterations in fat metabolism lead to the production and accumulation of ketones. This inappropriate persistence of the fasted state postprandially can be reversed by the administration of insulin.

The incidence of type 1 diabetes varies widely in different populations. Scandinavia and northern Europe have the highest incidence of type 1 diabetes: the yearly incidence per 100,000 youngsters 14 years of age or less is as high as 40 in Finland, 31 in Sweden, 22 in Norway, 27 in Scotland, and 20 in England. The incidence of type 1 diabetes generally decreases across the rest of Europe to 11 in Greece and 9 in France. Surprisingly, the island of Sardinia has as high an incidence as Finland, even though in the rest of Italy, including the island of Sicily, the incidence is only 11 per 100,000 per year. The United States averages 16 per 100,000. The lowest incidence of type 1 diabetes worldwide is less than 1 per 100,000 per year in China and parts of South America.

Worldwide incidence of type 1 diabetes continues to increase steadily. In Finland, the incidence has more than tripled since 1953, when it was 12/100,000/year, with an average increase of 2.4% per year. The EURODIAB study group reported recently 0.6% to 9.3% annual increases in incidence of type 1 diabetes in children younger than 15 years in various European countries. The most rapid increases have occurred in low-prevalence countries and in younger patients. Changes in environmental factors most likely explain this increased incidence.

Latent autoimmune diabetes of adulthood (LADA)   Type 1 diabetes can present at any age, although peaks in incidence occur before school age and around puberty. Older adults often present with a more indolent onset that sometimes leads to misdiagnosis and has led to the use of the term **latent autoimmune diabetes of adulthood (LADA)** to distinguish these patients. These initially unrecognized patients may retain enough β cell function at the outset to avoid ketosis, but develop increasing dependence on insulin therapy over time as their β cell mass diminishes. Islet cell antibody surveys among northern Europeans indicate that up to 15% of patients previously diagnosed with type 2 diabetes may actually have LADA.

Autoimmunity and Type 1 Diabetes

Most patients with type 1 diabetes at diagnosis have circulating antibodies against β cell proteins: islet cell antibodies (ICA), insulin autoantibodies (IAA), and antibodies to glutamic acid decarboxylase 65 (GAD), tyrosine phosphatase IA2 (ICA512), and zinc transporter 8 (ZnT8) (Table 17–6). These autoreactive antibodies can often be detected well before the onset of frank hyperglycemia, even decades earlier, providing evidence that the autoimmune process may be prolonged. After diagnosis, autoantibody levels often decline with increasing duration of the disease. Also, once patients are treated with insulin, low levels of IAA develop, even in patients that do not have an autoimmune etiology for their diabetes.

Although useful for diagnosing and predicting type 1 diabetes, antibodies against β cell proteins do not directly cause the

TABLE 17–6 Diagnostic sensitivity and specificity of autoimmune markers in newly diagnosed patients with type 1 diabetes mellitus.

	Sensitivity (%)	Specificity (%)
Glutamic acid decarboxylase (GAD65)	70-90	99
Insulin (IAA)	40-70	99
Tyrosine phosphatase IA2 (ICA512)	50-70	99
Zinc transporter 8 (ZnT8)	50-70	99

destruction of β cells in type 1 diabetes. Instead, it is the cellular immune system, the T lymphocytes that infiltrate the islets (a process called insulitis) and destroy the β cells. At the time of diagnosis, the islets of patients with type 1 diabetes are extensively infiltrated with T lymphocytes.

Normally, the thymus deletes autoreactive T cells during development so that the immune system becomes tolerant of self-antigens. In addition, certain specialized T cells, the regulatory T cells, further prevent attacks against healthy tissues by restraining the activity of any autoreactive T cells that escape the thymus. Type 1 diabetes results from a breakdown in these processes of self-tolerance in the immune system.

Type 1b diabetes   Approximately 5% of patients with the clinical features of type 1 diabetes lack serum evidence of autoimmunity. Some of these individuals have high-risk HLA haplotypes (discussed later) and may have T-cell–mediated β cell destruction in the absence of measurable levels of the known autoantibodies. Others in this group have low-risk HLA haplotypes, and appear to have a nonautoimmune cause for loss of β cell function. Such nonautoimmune type 1 diabetes has been referred to as type 1b diabetes, but a variety of terms has been used. This probably represents a heterogeneous group of disorders that lead to profound β cell dysfunction or loss, absolute insulin deficiency and a syndrome clinically similar to autoimmune type 1a diabetes. Under the accepted classification system, as specific disorders within this subgroup become defined and the genetic or environmental causes are identified, these disorders become reclassified within the monogenic or secondary forms of diabetes.

Included within this group are patients that present with a course of relapsing diabetic ketoacidosis with intervening normoglycemia that eventually progresses to permanent insulin-deficient diabetes. This disorder, ketosis prone diabetes (KPD, discussed later), has also been referred to as type 1b diabetes, and may result from unknown environmental insults combined with genetic defects in the β cell.

Autoimmune diabetes and stiff person syndrome   GAD antibodies, the first identified in type 1 diabetes, remain among the most clinically useful. Human pancreatic β cells produce GAD65, which functions as an enzyme that catalyzes the synthesis

of GABA from glutamate. GAD65 and the closely related isoform GAD67 are also found in central nervous system inhibitory neurons that secrete GABA. Some patients with GAD antibodies develop a rare neurologic condition, **stiff person syndrome**, caused by the depletion of GABA in the central nervous system and characterized by progressive rigidity and fluctuating muscle spasms. Approximately half of the patients with stiff person syndrome develop type 1 diabetes.

The vast majority of patients with type 1 diabetes do not develop symptoms of stiff person syndrome, despite the presence of GAD antibodies. The rare patients that develop the syndrome usually have much higher titers of GAD antibodies than typical patients with type 1 diabetes alone.

Genetics of Type 1 Diabetes

Family members of patients with type 1 diabetes have an increased lifetime risk of developing type 1 diabetes. The offspring of a mother with type 1 diabetes have a risk of 3%, whereas the risk is 6% for children of affected fathers. The risk in siblings of affected individuals is related to the number of human leukocyte antigen (HLA) haplotypes (discussed later) that the sibling shares. If one haplotype is shared, the risk is 6% and if two haplotypes are shared, the risk increases to 12% to 25%. For monozygotic twins, the concordance rate reaches 25% to 50%. Although these data demonstrate a strong genetic contribution to the risk of type 1 diabetes, genetics plays an even larger role in type 2 diabetes, and environment also clearly contributes substantially to the risk of type 1 diabetes.

Genes in the major histocompatibility (MHC) locus on the short arm of chromosome 6 explain at least half of the familial aggregation of type 1 diabetes. Within the MHC locus lie a number of closely packed genes involved in the function and regulation of the immune response. Although a number of genes within the MHC locus have been linked to the risk of developing type 1 diabetes, the most important of these are the genes encoding the HLA class II molecules DQ and DR. The professional antigen-presenting cells—dendritic cells, macrophages and B lymphocytes—use the class II molecules on their cell surface to present peptide antigens to T lymphocytes through the T-cell receptor. T cells activated by antigen-presenting cells carry out the β cell destruction that leads to type 1 diabetes. Although exact mechanisms remain uncertain, the variations in the amino acid sequence of individual HLA class II molecules may impact their ability to present specific self-peptides to T cells either in the process of central or peripheral tolerization or later during the development of the autoimmune response, thereby contributing to the risk of developing type 1 diabetes.

The DR haplotypes DR3 and DR4 are major susceptibility risk factors for type 1 diabetes. As many as 95% of type 1 diabetic patients have a DR3 or a DR4 haplotype—or both—compared with 45% to 50% of Caucasian nondiabetic controls. Individuals who express both a DR3 and a DR4 allele carry the highest risk for type 1 diabetes in the United States.

The high-risk *DR* genes are generally in linkage disequilibrium with *DQ* genes that themselves confer high risk, particularly DQA1*0501, DQB1*0201 (coupled with DR3), and DQA1*0301, DQB1*0302 (coupled with DR4). DQ alleles are associated not only with risk for type 1 diabetes but also with dominant protection, often in linkage with HLA-DR2. The most protective of these—and a quite common allele—is DQA1*0102, DQB1*0602. It occurs in over 20% of individuals in the United States but in less than 1% of children who develop type 1 diabetes.

An independent genetic link to chromosome 11 has also been identified in type 1 diabetes. Studies of a polymorphic DNA locus flanking the 5′ region of the *insulin* gene on chromosome 11 revealed a small but statistically significant linkage between type 1 diabetes and this genetic locus in a Caucasian population with type 1 diabetes. This polymorphic locus, which consists of a variable number of tandem repeats (VNTRs) with two common sizes in Caucasians, small (26-63 repeats) or large (140-243 repeats), does not encode a protein. An intriguing proposal to explain how the VNTR might influence susceptibility to type 1 diabetes was based on findings that *insulin* gene transcription is facilitated in the fetal thymus gland by the presence of the large allele of the VNTR locus flanking the *insulin* gene. The large VNTR allele might produce a dominant protective effect by promoting negative selection (deletion) by the thymus of insulin-specific T lymphocytes that play a critical role in the immune destruction of pancreatic β cells.

The established genetic association with the MHC region of chromosome 6 contributes much more (about 50%) to the genetic susceptibility to type 1 diabetes than does this locus flanking the *insulin* gene on chromosome 11, which contributes about 10%. Both candidate gene studies and genome-wide association studies (GWAS) have identified a number of additional risk loci that make smaller contributions to the genetic risk of type 1 diabetes. Many of the genes linked to these additional loci also play important roles in the function and regulation of the immune response.

Mutations in two genes involved in T-cell tolerance cause rare syndromes of monogenic autoimmune diabetes together with other autoimmune diseases. In the autosomal recessive disease autoimmune polyglandular syndrome type 1 (APS1; see Chapter 2), homozygous mutations in the gene encoding the autoimmune regulator (*AIRE*) prevent the expression of certain self-proteins in the thymus, thus allowing mature autoreactive T cells to leave the thymus. In addition to other autoimmune diseases and mucocutaneous candidiasis, approximately 20% of patients with APS1 develop type 1 diabetes. The second gene, *FOXP3*, found on the X chromosome, encodes a transcription factor required for the formation of regulatory T cells. Mutations in *FOXP3* cause immunodysregulation polyendocrinopathy enteropathy X-linked (IPEX) syndrome. IPEX presents in male patients with very early onset type 1 diabetes, often neonatal, combined with other autoimmune endocrinopathies, autoimmune skin disorders, diarrhea secondary to autoimmune enteropathy, and frequent severe infections.

Environmental Factors in Type 1 Diabetes

While genetic inheritance may play an important role in causing type 1 diabetes, the monozygotic twin studies demonstrate that

other causes, stochastic or environmental, are at least as important. Most individuals with type 1 diabetes do not have other family members with the disease. Environmental factors associated with increased risk of type 1 diabetes include viruses (mumps, congenital rubella, Coxsackie virus B4), toxic chemical agents such as vacor (a nitrophenylurea rat poison), and other destructive cytotoxins such as hydrogen cyanide from spoiled tapioca or cassava root. How these environmental insults lead to type 1 diabetes is unknown; they may directly damage β cells in some cases, or may act as initiators or accelerators of the autoimmune attack on the β cells. In some cases, molecular mimicry, wherein the immune system mistakenly targets β cell proteins that share homologies with certain viral or other foreign peptides may play a role.

Epidemiological studies have demonstrated an association between breast-feeding in the first 6 months of life and protection from type 1 diabetes. While it has been suggested that proteins in cow's milk may be the culprits, the strongest evidence supports the idea that human breast milk may reduce the risk of autoimmune disease.

Accumulating evidence shows that in the process of modernizing and improving public health, the risk of type 1 diabetes has increased, possibly due to the removal of some protective factors. Type 1 diabetes is almost unheard of in many third-world countries, and has its highest incidence in countries with the best public health systems, such as the Scandinavian countries. In addition, the incidence of the disease has been steadily increasing over the past century in western and westernizing countries and is especially high among the more affluent. This has led to the suggestion that a dirty environment, one with more infections (especially more parasitic diseases) and more antigen exposure, may reduce the risk of type 1 disease.

TYPE 2 DIABETES

Type 2 diabetes mellitus—previously called non-insulin-dependent diabetes or adult-onset diabetes mellitus—results from relative insulin deficiency, in contrast to the absolute insulin deficiency of patients with type 1 diabetes. Type 2 diabetes is a heterogeneous disorder and probably represents a large number of different primary genetic and environmental insults leading to relative insulin deficiency—a mismatch between insulin production and insulin requirements. Clinically, patients with type 2 diabetes can range from those with severe insulin resistance and minimal insulin secretory defects to those with a primary defect in insulin secretion.

Type 2 diabetes accounts for 80% to 90% of cases of diabetes in the United States. These patients commonly present as adults with some degree of obesity, although increasing rates of obesity are leading to earlier onset of the disease in adolescents and children. At onset, most patients with type 2 diabetes do not require insulin to survive, but over time their insulin secretory capacity tends to deteriorate, and many eventually need insulin treatment to achieve optimal glucose control. Ketosis seldom occurs spontaneously, and if present, it is a consequence of severe stress from trauma or infection.

TABLE 17–7 Factors reducing sensitivity to insulin.

Prereceptor	Insulin autoantibodies Reduced transendothelial transit
Primary defect in insulin signaling	Insulin receptor mutations Leprechaunism (complete) Rabson-Mendenhall syndrome (partial) Type A (mild) Defects in other genes involved in insulin signaling Insulin receptor autoantibodies (Type B) Ataxia telangiectasia syndrome
Secondary to other endocrine disorders	Cushing syndrome Acromegaly Pheochromocytoma Glucagonoma Hyperthyroidism Insulinoma
Secondary to other disorders	Visceral obesity Stress (infection, surgery, etc) Uremia Hyperglycemia (mild resistance seen in type 1 diabetes) Liver disease Cytogenetic disorders (Down, Turner, Klinefelter) Neuromuscular disorders (muscular dystrophies, ataxias, muscle inactivity) Congenital lipodystrophies/lipoatrophy Acquired lipodystrophy
Secondary to normal physiologic states	Puberty Pregnancy Starvation
Secondary to medications	Glucocorticoids Atypical antipsychotic drugs Antiretroviral protease inhibitors Nicotinic acid Thiazide diuretics Oral contraceptive Progesterone β blockers

Most patients with type 2 diabetes, irrespective of weight, have some degree of tissue insensitivity to insulin attributable to several interrelated factors (Table 17–7). These include putative (mostly as yet undefined) genetic factors, which are aggravated in time by further enhancers of insulin resistance such as aging, a sedentary lifestyle, and abdominal visceral obesity. Not all patients with obesity and insulin resistance develop hyperglycemia, however. An underlying defect in the ability of the β cells to compensate for the increased demand determines which patients will develop diabetes in the setting of insulin resistance. Furthermore, both the tissue resistance to insulin and the impaired β cell response to glucose appear to be further aggravated by sustained hyperglycemia, which may impede both insulin signaling and β cell function. Treatment that reduces the blood glucose levels toward normal reduces this acquired defect in insulin resistance and may also improve glucose-induced insulin release to some degree, although the long-term decline in β cell function continues.

Type 2 diabetes frequently goes undiagnosed for many years because the hyperglycemia may develop quite gradually and without initial symptoms. Despite this mild presentation, these patients develop microvascular, and, especially, macrovascular complications. Furthermore, as noted earlier, patients with type 2 diabetes suffer from a progressive decline in β cell capacity, leading to worsening hyperglycemia over time.

A. Obesity in type 2 diabetes The majority of people with type 2 diabetes have excess adiposity, although the prevalence of obesity in association with type 2 diabetes varies among different racial groups. Sixty to eighty percent of North Americans, Europeans, or Africans with type 2 diabetes and close to 100% of individuals with type 2 disease among Pima Indians or Pacific Islanders from Nauru or Samoa have obesity as defined by body mass index (BMI, see Chapter 20), while as few as 30% of Chinese and Japanese patients with type 2 diabetes are obese. However, many of those individuals with type 2 diabetes who do not meet BMI criteria for obesity have a predominantly abdominal distribution of fat, producing an abnormally high waist to hip ratio. Increases in visceral adiposity correlate with increased insulin resistance.

B. Insulin resistance in type 2 diabetes Insulin resistance can be broadly defined as a decrease in tissue responsiveness to insulin. Clinically it can be assessed directly by measuring the ability of a fixed dose of insulin to promote total body glucose disposal. It can be assessed indirectly by measuring fasting insulin levels. An increase in insulin levels with normal plasma glucose indicates insulin resistance.

As adiposity increases, especially abdominal visceral fat deposits, total body insulin sensitivity decreases. Since adipose tissue only removes a small fraction of plasma glucose, clearly the increased adipose fat stores impact total body insulin sensitivity through effects on other tissues, especially muscle and liver, causing them to decrease insulin-stimulated glucose disposal. The exact means by which fat storage in adipocytes affects the insulin sensitivity of other cells remains uncertain, but experimental evidence suggests several possible mechanisms.

Abnormalities of insulin receptors—in concentration, affinity, or both—affect insulin action. Target tissues downregulate the number of insulin receptors on the cell surface in response to chronically elevated circulating insulin levels, probably by increased intracellular degradation. When insulin levels are low, on the other hand, receptor binding is upregulated. Conditions associated with high insulin levels and lowered insulin binding to the receptor include obesity, high intake of carbohydrates, and chronic exogenous overinsulinization. Conditions associated with low insulin levels and increased insulin binding include exercise and fasting. The insulin receptor itself is probably not the major determinant of insulin sensitivity under most circumstances, however. Clinically relevant insulin resistance most commonly results from defects in postreceptor intracellular signaling pathways.

Adipokines Adipose tissue can affect the insulin sensitivity of other tissues through the secretion of signaling molecules, *adipokines*, that inhibit (TNF-α, IL-6, leptin, resistin, and others) or enhance (adiponectin) insulin signaling locally or in distal target tissues (see Chapter 20). Levels of fat storage in adipocytes, along with insulin signaling itself, regulate the production and secretion of many of the adipokines. Some of these mediators of insulin resistance may reduce insulin signaling by blocking access of insulin to target tissues through reduced transendothelial transit. However, most evidence suggests that the secreted adipokines influence insulin signaling in distant tissues through effects on postreceptor intracellular signaling pathways. Potential intracellular effectors include protein tyrosine phosphatases that dephosphorylate the receptor and pathway components, inhibitors such as the SOCS proteins that block receptor–IRS interactions, and serine/threonine kinases that inhibit the receptor and substrates through serine phosphorylation.

Free fatty acids and ectopic lipid storage The release of fatty acids by the engorged adipocytes (especially visceral adipocytes, from which fatty acids are more readily mobilized) may play a role in the development of insulin resistance as well. Oxidation of fatty acids by muscle and other tissues could inhibit glycolysis and reduce insulin-stimulated glucose removal (the Randle hypothesis, named after its original proponent). Increased fat storage in adipocytes and release of fatty acids may also eventually cause a shift in lipid storage, increasing lipid uptake and storage in nonadipose tissues such as muscle, liver, and β cells. Ectopic lipid storage in these tissues may lead to a decrease in their insulin sensitivity. In addition, free fatty acids may function directly in a signaling role both locally within the adipose tissue and systemically.

Inflammation In addition to adipocytes, adipose tissue contains a variety of other cell types including inflammatory/immune cells, such as macrophages and lymphocytes. Recent evidence implicates these cells in obesity-induced insulin resistance. As adipocyte lipid stores rise, the increased release of free fatty acids and proinflammatory adipokines recruits macrophages to the adipose tissue and activates them. The activated macrophages then release a variety of molecules (TNF-α, IL-6, nitric oxide, and others) that decrease the insulin sensitivity of the adipocytes and further increase their release of proinflammatory fatty acids and peptides, creating a positive feedback loop that maintains a chronic state of local inflammation and insulin resistance. Release of these adipokines and proinflammatory cytokines, along with the increased release of free fatty acids and the development of ectopic lipid accumulation, promotes the development of inflammation and insulin resistance in the other key insulin-target tissues, such as muscle and liver. Similar mechanisms could also lead to inflammation in the islets and contribute to β cell failure.

PPARγ activity in the adipose tissue generally has beneficial effects on systemic insulin signaling through several mechanisms: (1) Promotion of adipose lipid storage, which thereby decreases ectopic lipid storage in nonadipose tissues; (2) Inhibition of the production of adipokines and proinflammatory cytokines, which promote insulin resistance, by adipocytes; (3) Promotion of the alternative activation of macrophages to the anti-inflammatory

M2 state, rather than the proinflammatory M1 state; (4) Inhibition of the release of proinflammatory and proresistance cytokines by macrophages. Although it is expressed at much lower levels in muscle than in adipose tissue, PPARγ in myocytes might also have a direct role in controlling muscle insulin sensitivity; however, these findings remain controversial.

Tissue heterogeneity in insulin resistance   Finally, it must be kept in mind that not all tissues necessarily develop insulin resistance in parallel, and in many cases insulin resistance may be limited to specific tissues. The combination of local and systemic contributors to obesity-induced insulin resistance may explain the different levels of insulin resistance in different tissues of the same patient. Even in the same cell, insulin resistance may impact different arms of the insulin-signaling pathway discordantly. As insulin levels rise in response to insulin resistance, heterogeneity in insulin sensitivity could explain unique syndromes associated with insulin resistance like hepatic steatosis and polycystic ovary syndrome.

Other causes of insulin resistance   Visceral obesity is not the only cause of insulin resistance, although it is by far the most common cause in most populations. Other causes of insulin resistance include a variety of genetic and acquired defects that impact the insulin receptors or postreceptor signaling pathways (see Table 17–7).

Clinical consequences of insulin resistance   In addition to the impact on glucose metabolism, severe insulin resistance and the resulting elevation in circulating insulin levels can cause other clinical consequences including acanthosis nigricans, pseudoacromegaly, and hyperandrogenism. Acanthosis nigricans appears to be a consequence of very high circulating insulin levels that cross over to bind to IGF receptors on epidermal and melanin-containing cutaneous cells. This leads to local skin hyperplasia with papillomatosis, hyperkeratosis, and hyperpigmentation. The dark, velvety patches of skin most commonly appear on the back of the neck, axillae, and anticubital fossae. In extreme and prolonged cases of insulin resistance, the secondary increase in signaling through the IGF-1 receptor, or possibly residual signaling through the mitogenic arm of the insulin signaling pathway, can cause pseudoacromegaly, a syndrome with all of the bone and soft tissue changes of acromegaly (see Chapter 4), but no elevation in growth hormone or IGF-1. A similar action of extremely high insulin levels on ovarian hilar cells has been implicated in women with insulin resistance who develop hyperandrogenism and hirsutism associated with menstrual irregularities, enlarged cystic ovaries and infertility (polycystic ovary syndrome).

C. β Cell defects in type 2 diabetes   Although the majority of people with type 2 diabetes have insulin resistance, most people with insulin resistance do not have diabetes because their β cells compensate for the insulin resistance by producing and secreting more insulin. Those individuals with insulin resistance who develop type 2 diabetes have a defect in the compensatory response of their β cells to insulin resistance. Functionally, this defect is revealed by a reduction in first phase insulin secretion and the maximal insulin secretion stimulated by glucose.

While increased insulin secretion per β cell may contribute to the compensatory response to insulin resistance, increases in the number of β cells play a role as well. In the setting of obesity, hyperplasia of pancreatic β cells is often present and probably accounts for the normal or exaggerated insulin responses to glucose and other stimuli seen in obese individuals without type 2 diabetes. Assessment of total β cell mass at autopsy has revealed that β cell mass increases in obesity, but the individuals with type 2 diabetes have decreased β cell mass when compared to nondiabetic individuals with the same BMI.

Several possible defects could contribute to the failure of β cell mass compensation in people with type 2 diabetes. Underlying genetic differences in the pathways that drive β cell expansion appear to limit compensation in individuals with high genetic risk of diabetes. In susceptible individuals with obesity, ectopic fat deposition in the islets, local obesity-induced inflammation in the islets, and local and circulating adipokines and inflammatory cytokines may accelerate β cell loss. As β cell failure progresses, levels of glucose and free fatty acids start to rise, which in turn can cause further β cell toxicity. Increased demand on a decreased β cell mass may cause further damage through ER stress and the increased formation of toxic IAPP oligomers. Then, once diabetes is established, all of these mechanisms may further contribute to the progressive decline in β cell function that characterizes type 2 diabetes.

D. Metabolic syndrome   Patients with visceral obesity and insulin resistance often present with a cluster of abnormalities commonly termed the metabolic syndrome. Hyperglycemia in these patients is frequently associated with hyperinsulinemia, dyslipidemia, and hypertension, which together lead to coronary artery disease and stroke. It has been suggested that this aggregation results from a genetic defect producing insulin resistance, particularly when obesity aggravates the degree of insulin resistance. In this model, impaired action of insulin predisposes to hyperglycemia, which in turn induces hyperinsulinemia. If this hyperinsulinemia is of insufficient magnitude to correct the hyperglycemia, type 2 diabetes is manifested. The excessive insulin level could also increase sodium retention by renal tubules, thereby contributing to or causing hypertension. Increased VLDL production in the liver, leading to hypertriglyceridemia (and consequently a decreased high-density lipoprotein [HDL] cholesterol level), has also been attributed to hyperinsulinism. Moreover, it has been proposed that high insulin levels can stimulate endothelial and vascular smooth muscle cell proliferation—by virtue of the hormone's action on growth factor receptors—to promote atherosclerosis.

Although there is full agreement on an association of these disorders, the mechanism of their interrelationship remains speculative and open to experimental investigation. Controversy persists about whether or not hypertension is caused by the hyperinsulinism that results from insulin resistance. Moreover, patients with hyperinsulinism due to an insulinoma are generally normotensive,

and there is no reduction of blood pressure after surgical removal of the insulinoma restores normal insulin levels.

An alternative unifying hypothesis could be that visceral obesity directly induces the other components of this syndrome. Visceral obesity is an independent risk factor for all of the other components of the metabolic syndrome. In addition to the metabolic effects of visceral obesity, the adipokines and inflammatory cytokines generated from overloaded and inflamed adipose tissue may contribute to the pathophysiology of the syndrome. Although the full details of the role of these molecules in causation of the metabolic syndrome remain under investigation, the adipocytes and associated macrophages clearly are not just innocent bystanders but play active roles in the development of systemic insulin resistance, hypertension, and hyperlipidemia. Furthermore, thrombi in atheromatous vessels may be more hazardous in patients with visceral obesity because of an associated increase in plasminogen activator inhibitor-1 (PAI-1), a circulating factor produced by omental and visceral adipocytes that inhibits clot lysis. This model emphasizes the importance of measures such as diet and exercise that reduce visceral adiposity in the management of patients with metabolic syndrome and obese type 2 diabetes.

The main value of grouping these disorders as a syndrome, regardless of its nomenclature, is to remind physicians that the therapeutic goals in these patients must not only correct hyperglycemia but also manage the elevated blood pressure and hyperlipidemia that result in considerable cardiovascular morbidity as well as cardiovascular deaths. In addition, it reminds physicians that when choosing antihypertensive agents or lipid-lowering drugs to manage one of the components of this syndrome, their possible untoward effects on other components of the syndrome should be carefully considered. For example, physicians aware of this syndrome are less likely to prescribe antihypertensive drugs that raise lipids (diuretics, beta blockers) or that raise blood glucose (diuretics). Likewise, they may refrain from prescribing drugs that correct hyperlipidemia, but increase insulin resistance with aggravation of hyperglycemia (nicotinic acid).

E. Genetics of type 2 diabetes   Type 2 diabetes has a strong genetic link. Depending on the population studied, monozygotic twins have lifetime concordance rates for type 2 diabetes exceeding 90%. In contrast, concordance rates for type 1 diabetes in monozygotic twins are 25% to 50%. Most individuals with type 2 diabetes have other family members with the disease, but the inheritance rarely fits Mendelian patterns, supporting the conclusion that multiple genes with varying degrees of penetrance contribute. Because of the heterogeneous nature of type 2 diabetes, and its complex inheritance, efforts to identify the genes that contribute to the disease have had very limited success in the vast majority of affected patients. There has been considerable success, however, in identifying small subsets of patients with unique monogenic forms of the disease. When the etiologic defect has been defined, these patients have been reclassified within a group designated "Monogenic Diabetes" (see Table 17–5).

Efforts to identify the genes involved in polygenic type 2 diabetes have focused on two approaches: candidate gene testing and GWAS. To date candidate gene and GWAS approaches have identified 19 loci with common variants linked to type 2 diabetes. Although statistically significant and validated in additional populations, these loci independently make very small contributions to type 2 diabetes risk. The highest risk of these common variants is found at a locus adjacent to the gene encoding TCF7L2, a transcription factor involved in Wnt signaling and implicated in β cell turnover. Inheritance of the high-risk TCF7L2 allele increases the probability of developing diabetes by 1.5-fold.

Among the genes identified so far, most are involved in β cell function and turnover. When combined with the predominance of β cell genes implicated in mongenic forms of diabetes (see Table 17-5), these results reinforce the critical role of the β cell in controlling blood glucose and its involvement in the pathophysiology of type 2 diabetes. Hopefully, with the advent of high-throughput whole genome sequencing technologies, the identification of rarer, but higher risk, variants will further expand our understanding of the genetics of type 2 diabetes in the near future.

F. Environmental factors in type 2 diabetes   Despite the critical role of genetics in type 2 diabetes, environment contributes as well, especially in determining the age of onset and severity of the disease. There is generally a low incidence of type 2 diabetes in underdeveloped countries, especially in rural areas. Western countries and westernizing countries suffer from a much higher incidence. Over the past half-century, the incidence of type 2 diabetes has increased rapidly in almost all world populations but especially in emerging third-world countries. This increase correlates with increasing rates of obesity in the same populations and reflects increased access to food with high caloric content and decreased physical activity. This combination inevitably leads to increased adiposity, especially in the more readily mobilized fat stores surrounding the viscera in the abdomen.

One of the most dramatic recent changes in the epidemiology of diabetes has been the growing incidence of type 2 diabetes in children. While rarely seen in children a generation ago, type 2 diabetes is now as common as type 1 diabetes in teenagers in the United States and is seen with increasing frequency, even in younger children. Again, this increase is directly related to increasing visceral adiposity.

MONOGENIC DIABETES

Autosomal Dominant Genetic Defects of Pancreatic β Cells

MODY   This subgroup of monogenic disorders is characterized by the onset of diabetes in late childhood or before the age of 25 years as a result of a partial defect in glucose-induced insulin release and accounts for up to 5% of diabetes in North American and European populations. A strong family history of early-onset diabetes occurring in one parent and in one-half of the parent's offspring suggests autosomal dominant transmission. In contrast to most patients with type 2 diabetes, these patients are generally nonobese and lack

associated insulin resistance. Instead they exhibit predominantly a defect in glucose-stimulated insulin secretion. However, because they are not ketosis-prone and may initially achieve good glycemic control without insulin therapy, their disease has been called **maturity-onset diabetes of the young (MODY)**. Several distinct types have been described with single-gene defects, and all have been shown to produce a defect in glucose-induced insulin release. MODY 2 results from an abnormal glucokinase enzyme. Many of the other forms of MODY are due to mutations of nuclear transcription factors that control β cell function and formation by regulating the expression of genes in β cells or their developmental precursors (Figure 17–8 and Table 17–11).

MODY 1 includes multiple members of a large pedigree known as the R-W family, descendants of a German couple who immigrated to Michigan in 1861. They were studied prospectively since 1958, and in 1996 the genetic defect was shown to be a nonsense mutation of a nuclear transcription factor found in liver as well as in pancreatic β cells. This gene has been termed *hepatocyte nuclear factor 4α* (*HNF4α*) and is found on chromosome 20. Mutations of this gene are among the rarest of the MODY groups, with few reported in families outside the Michigan pedigree. These patients display a progressive decline in β cell function and eventually develop chronic complications of diabetes including microangiopathy at a rate approaching that of people with type 1 diabetes. They often fare better with insulin therapy.

MODY 2 was first described in French families but has now been found in racial groups from most parts of the world. Multiple different mutations of the *glucokinase* gene (*GCK*) on chromosome 7 have been identified and characterized. In pancreatic β cells, glucokinase controls the rate-limiting step in glycolysis and thereby determines the rate of ATP production from glucose and the insulin secretory response to glucose (see Figure 17–5). Reduced glucokinase activity resets the sensitivity of the β cell to

glucose so that it requires higher plasma glucose levels to stimulate insulin secretion, resulting in fasting hyperglycemia and mild diabetes. Although some of these mutations can completely block the enzyme's function, others interfere only slightly with its action. In contrast to all the other forms of MODY, most patients with one mutated *GCK* allele (heterozygotes) have a benign course with few or no chronic complications and respond well to diet therapy or oral antidiabetic drugs without the need for insulin treatment. On the other hand, rare individuals who inherit two mutated *GCK* alleles have permanent neonatal diabetes (discussed later), a nonimmune form of absolute insulin deficiency that presents at birth.

In contrast to the mutations that reduce glucokinase enzyme activity and cause MODY 2, rare mutations in *GCK* that increase the enzymatic activity of glucokinase can cause increased insulin secretion and hypoglycemia (see Chapter 18), demonstrating the key role of this enzyme in determining the sensitivity of the β cell to glucose.

MODY 3 is caused by mutations of *hepatocyte nuclear factor 1α* (*HNF1α*), whose gene is located on chromosome 12. This is the most common form of MODY in European populations, with many different mutations having been reported. Like *HNF4α*, the *HNF1α* transcription factor is expressed in pancreatic β cells as well as in liver. Also similar to *HNF4α*, mutations in *HNF1α* cause a progressive form of diabetes with declining β cell function that often leads to dependence on insulin therapy and high rates of microvascular complications. Patients carrying *HNF1α* mutations often have reduced circulating levels of C-reactive protein. Of note, early in the course of the disease, these patients may display an exaggerated response to sulfonylureas.

Together, *HNF1α* and *HNF4α*, along with several other β cell transcription factors including *PDX1* and *NEUROD1* (discussed later), form an interacting network of transcription factors (Figure 17–11). This transcriptional network regulates
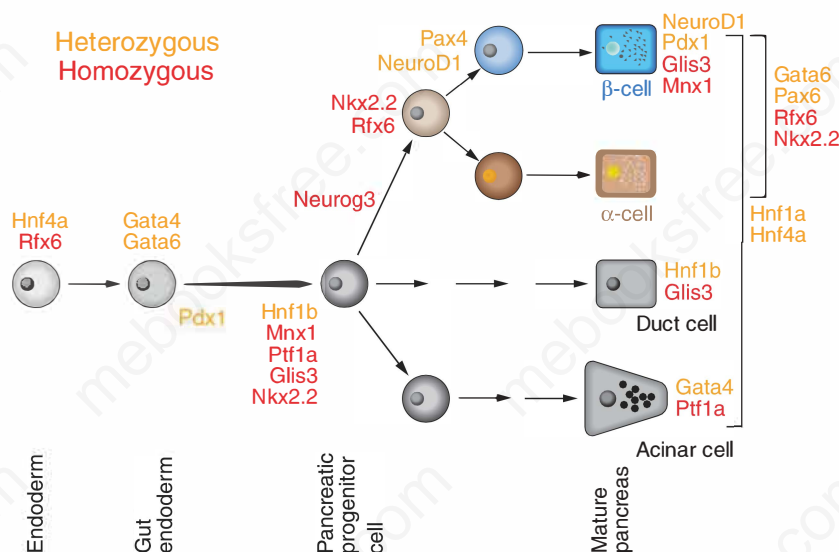
FIGURE 17–11 A simplified model is shown for the differentiation of pancreatic cell types during the development of the pancreas, starting with the formation of the definitive endoderm. Transcription factors involved in this process and also implicated in human diabetes are shown at their major expression time points, and are labeled in orange if heterozygous mutations cause diabetes and in red if only homozygous mutations have been shown to cause diabetes.

genes involved in both the development of the β cell as well as mature β cell function, including glucose-sensing and insulin secretion. Target genes include *GCK*, as well as genes implicated in the formation, maturation and expansion of β cells. Impairment in β cell formation and regeneration may explain the progressive nature of the MODY transcription factor syndromes and reinforces the importance of β cell mass in preventing hyperglycemia.

MODY 4 results from mutation of a pancreatic nuclear transcription factor known as *pancreatic and duodenal homeobox-1* (*PDX1*), whose gene is on chromosome 13. It mediates *insulin* gene transcription and regulates expression of other β cell-specific genes including *GCK*. When both alleles of this gene are nonfunctioning, agenesis of the entire pancreas results; but in the presence of a heterozygous mutation of *PDX1*, a mild form of MODY has been described in which affected individuals developed diabetes at a later age (mean onset at 35 years) than occurs with the other forms of MODY, in whom onset generally occurs before the age of 25 years.

MODY 5 was initially reported in a Japanese family with a mutation of *HNF1*β, a hepatic nuclear transcription factor closely related in structure and molecular function to *HNF1*α. The two *HNF1* factors, however, are expressed by different cells. *HNF1*β is expressed early in the development of the liver, pancreas, kidneys, and genitourinary system, and is not found in mature β cells.

Mutations in this gene cause a moderately severe form of MODY with progression to insulin treatment and severe diabetic complications in those affected. Consistent with its expression pattern early in development, *HNF1*β mutations also frequently cause reduction in the overall size of the pancreas, decreased insulin production, and congenital defects in the kidney and urogenital tract. Patients may also suffer from varying degrees of cholestatic jaundice, hyperuricemia, nephropathy, and hypomagnesemia secondary to renal magnesium wasting.

MODY 6, a milder form of MODY similar to MODY 4, results from mutations in the gene encoding the islet transcription factor NeuroD1. Like PDX1, NeuroD1 plays an important role in the expression of insulin and other β *cell* genes, and in the formation and maintenance of β cells.

Other MODY genes   The six *MODY* genes listed earlier explain the majority of cases of MODY in patients of European ancestry, but less than half of those in non-European populations. Several rare variants in other genes have been implicated in autosomal dominant diabetes in a few families (see Table 17–8); however, a consensus has not yet been reached that these families fit the criteria for MODY and that the reported variants cause the disorder. The causative genes in most non-Europeans with MODY remain unknown.

The identification of mutations in multiple genes encoding pancreatic transcription factors in patients with MODY has led to the screening of other genes encoding pancreatic transcription factors in patients with diabetes. Heterozygous mutations in genes encoding several transcription factors have been identified in patients with earlier or later onset diabetes than in MODY patients, including *GATA6* and *GATA4* (neonatal diabetes) and

TABLE 17–8 Autosomal dominant genetic defects of pancreatic β cell function.

Syndrome	Mutated Protein	Gene	Function
MODY 1	Hepatocyte nuclear factor-4α	*HNF4A*	Transcription factor
MODY 2	Glucokinase	*GCK*	Glucose sensor
MODY 3	Hepatocyte nuclear factor-1α	*HNF1A*	Transcription factor
MODY 4	Pancreatic duodenal homeobox-1	*PDX1*	Transcription factor
MODY 5	Hepatocyte nuclear factor-1β	*HNF1B*	Transcription factor
MODY 6	NeuroD1	*NEUOROD1*	Transcription factor
MODY 7	Kruppel-like factor 11	*KLF11*	Transcription factor
MODY 8	Carboxyl-ester lipase	*CEL*	Exocrine enzyme
MODY 9	Paired homeobox 4	*PAX4*	Transcription factor
MODY 11	Tyrosine-protein kinase Blk	*BLK*	Intracellular signaling
MODY 14	Adaptor Protein APPL1	*APPL1*	Intracellular signaling
GATA4	GATA-binding protein 4	*GATA4*	Transcription factor
GATA6	GATA-binding protein 6	*GATA6*	Transcription factor
Mutant insulin or Proinsulin	Preproinsulin	*INS*	Hormone
KATP mutations	Inward-rectifying K⁺ channel 6.2	*KCNJ11*	Ion channel
	Sulfonylurea receptor 1	*ABBC8*	Ion channel

ISL1, *PAX6*, and *PAX4* (later onset diabetes). Heterozygous GATA4 and GATA6 mutations can also be associated with pancreatic agenesis and congenital cardiac defects. The association of diabetes with heterozygous mutations in so many β cell genes highlights the critical importance of optimal β cell function in metabolic regulation. Even modest defects in glucose-induced insulin secretion can result in hyperglycemia.

Insulin mutations   Sequencing of the *insulin* gene more than 30 years ago led to the first descriptions of heterozygous mutations in the coding sequence of the *insulin* gene that produce abnormal circulating forms of insulin. Most of these initial cases presented with high circulating levels of insulin, but normal insulin sensitivity and normal glucose levels. Because the abnormal insulins in these cases bind to receptors poorly, they have very low biologic activity and are cleared at a slower rate, leading to accumulation in the blood at higher levels than normal insulin and a subnormal

molar ratio of C peptide to immunoreactive insulin. They typically are not associated with hyperglycemia.

However, a mutation in the insulin B chain in one family was associated with decreased circulating levels of both the mutant and normal insulins, and diabetes. Subsequent extensive sequencing of the *INS* gene has identified many other mutant insulins that produce a similar heterozygous form of diabetes. Modeling of these dominant insulin mutations in mice demonstrates that they lead to the accumulation of aberrantly folded proteins in the endoplasmic reticulum, activation of the unfolded protein response in the ER, and β cell apoptosis. Patients with diabetes secondary to *INS* gene mutations usually present at a younger age than most patients with MODY, often developing the disease as neonates (see Neonatal Diabetes later). Because of the profound β cell loss, these patients follow a disease course similar to type 1 diabetes with absolute insulin deficiency and ketosis, and they require insulin therapy.

This syndrome highlights the sensitivity of the β cells to ER stress, which may explain why β cells often fail when presented with the increased insulin demands associated with insulin resistance, the toxicity of IAPP oligomers, or mutations in genes involved in the unfolded protein response pathway (discussed later).

Mutations in the subunits of the ATP-sensitive potassium channel

β cells sense rising blood glucose concentrations by increasing the production of ATP from glucose. The rising intracellular ATP levels cause the closure of ATP-sensitive potassium channels on the cell surface, which depolarizes the cell and sets off a cascade of events that leads to the secretion of insulin (see Figure 17–5). Rare dominant activating mutations in either of the two units of the channel, SUR1 and Kir6.2 (gene names *ABCC8* and *KCNJ11*, respectively), can cause the channels to remain open and prevent glucose-induced depolarization and insulin secretion. Children heterozygous for these mutations present with early-onset diabetes, commonly as neonates, and may have associated neurologic deficits implying a role for these channels in the central nervous system. Depending on the exact mutation, some of these children can still respond to treatment with sulfonylureas, which may also ameliorate the neurologic symptoms.

Other Genetic Defects of Pancreatic β Cells

Autosomal recessive genetic defects Although less common than the autosomal dominant β cell disorders, mutations in several genes causing autosomal recessive syndromes with defects in β cell function have been identified in patients with diabetes (Table 17–9). Due to the severity of the β cell defect, many of these present with neonatal diabetes. This group of disorders includes homozygous mutations in the MODY genes *GCK*, *PDX1,* and *NEUROD1*. Homozygous mutations in *GCK* cause a much more severe syndrome than the mild glucose-sensing defect seen in MODY 2. Patients with homozygous *GCK* mutations present at birth with severe hyperglycemia and require insulin therapy.

In patients with mutations in both alleles of *PDX1*, the pancreas fails to form, and they have pancreatic exocrine deficiency as well as diabetes. Homozygous mutations in several other pancreatic transcription factor genes have been described as well, including *PTF1A, NEUROG3, RFX6, NKX2-2, MNX1,* and *GLIS3*. Like *PDX1*, homozygous mutation of *PTF1A* causes diabetes and pancreatic agenesis, but it causes cerebellar atrophy as well. The transcription factor Neurog3 drives the formation of the endocrine cells in both the pancreas and gut. In addition to diabetes onset prior to puberty, infants born with homozygous *NEUROG3*

TABLE 17–9 Autosomal recessive genetic defects of pancreatic β cell function.

Function	Protein	Gene	Associated Defects
Glucose sensing	Glucokinase	GCK	
Transcription factors	Pancreatic duodenal homeodomain protein 1	PDX1	Pancreatic agenesis
	Pancreatic transcription factor 1a	PTF1A	Pancreatic and cerebellar agenesis
	Neurogenenin3	NEUROG3	Absence of gut endocrine cells, malabsorption
	Regulatory factor X-box 6	RFX6	Mitchell-Riley syndrome: absence of gut endocrine and islet α, β, and δ cells, pancreatic and gall bladder hypoplasia, intestinal atresia, malabsorption
	GLI Similar-3	GLIS3	Hypothyroidism, cholestasis, polycystic kidneys, hypoplastic pancreas
	NK homeodomain protein 2-2	NKX2-2	Severe developmental delay, hypotonia, cortical blindness, impaired visual tracking, and hearing impairment
	Motor neuron and pancreas homeodomain protein 1	MNX1	Severe developmental delay and neurological deficits, sacral agenesis
Unfolded protein response	Eukaryotic translation initiation factor 2-α kinase 3/PKR-like ER kinase (PERK)	EIF2AK3	Wolcott-Rallison syndrome: epiphyseal dysplasia and growth retardation; variable hepatic, renal, cardiac, and pancreatic exocrine defects
	Immediate early response 3 interacting protein 1	IER3IP1	MEDS syndrome: microcephaly, epilepsy
	Wolfram syndrome protein 1 (WFS1)	WFS1	Wolfram syndrome: diabetes insipidus, diabetes mellitus, optic atrophy, and deafness (DIDMOAD)
Thiamine transport	Solute carrier family 19 (thiamine transporter), member 2	SLC19A2	Thiamine-responsive megaloblastic anemia: anemia, deafness

mutations have severe malabsorption associated with a lack of gut endocrine cells from birth.

Homozygous mutations in *RFX6*, which encodes a transcription factor that functions downstream of *NEUROG3* and upstream of *PDX1* in β cell development, cause **Mitchell–Riley syndrome** in which neonates present with diabetes in association with complete absence of all islet cell types except PP cells, hypoplasia of the pancreas and gall bladder, intestinal atresia, and severe malabsorption.

The zinc finger transcription factor GLIS3 is expressed broadly in many tissues and plays a role in the transcription of the *insulin* gene. Homozygous mutations in *GLIS3* cause congenital hypothyroidism in addition to neonatal diabetes.

In the autosomal recessive **Wolcott–Rallison syndrome**, affected children present with neonatal diabetes, epiphyseal dysplasia, and growth retardation together with a variety of progressive hepatic, renal, cardiac, and pancreatic exocrine defects and developmental delay. The causative gene, *EIF2AK3*, encodes a kinase (PKR-like ER kinase [PERK]) activated by the presence of unfolded proteins in the ER. PERK controls one of the three parallel arms of the unfolded protein response; inositol-requiring enzyme 1 (IRE1) and activating transcription factor 6 (ATF6) activate the other two arms. Together these three molecules activate signaling pathways that protect the cell from ER stress but lead to apoptosis when these protective mechanisms fail. Mice lacking PERK have an inadequate response to ER stress, which leads to accelerated β cell apoptosis. With the high load of insulin production in the ER, β cells are uniquely sensitive to ER stress, and this sensitivity probably underlies the damage caused by mutant insulins, IAPP oligomers, and Wolfram syndrome as well.

Wolfram syndrome is an autosomal recessive neurodegenerative disorder first evident in childhood. Patients present with diabetes insipidus, diabetes mellitus, optic atrophy, and deafness—hence the acronym DIDMOAD. Diabetes mellitus usually develops in the first decade together with the optic atrophy, followed by central diabetes insipidus and sensorineural deafness during the second decade in 60% to 75% of patients. Ureterohydronephrosis, neurogenic bladder, cerebellar ataxia, peripheral neuropathy, and psychiatric illness develop later in many patients. The diabetes mellitus is nonimmune and not linked to specific HLA haplotypes, but on autopsy these patients have selective loss of β cells in the pancreas. Genetic studies mapped the causative mutations to a gene called *WFS1*, which encodes a 100.3-kDa transmembrane protein localized to the endoplasmic reticulum membranes of all cells. *WFS1* is expressed at particularly high levels in β cells. Studies in mice have shown that the WFS1 protein forms part of the unfolded protein response downstream of PERK and IRE1 and helps protect the β cells from ER stress and apoptosis, especially during periods of high insulin demand.

Children with **thiamine-responsive megaloblastic anemia syndrome** carry mutations in the high-affinity thiamine transporter SLC19A2 found on cell and mitochondrial membranes. They develop megaloblastic anemia, diabetes, and sensorineuronal deafness. The diabetes usually presents in the first decade of life. In the absence of SLC19A2, cells and mitochondria can still transport thiamine through lower affinity transporters, and both the anemia and the diabetes respond to pharmacologic treatment with thiamine. However, all patients eventually require insulin replacement despite thiamine therapy. It remains unclear how partial defects in cellular and mitochondrial thiamine transport cause β cell failure.

Mitochondrial DNA mutations   Because sperm do not contain mitochondria, only the mother transmits mitochondrial genes to her offspring. Diabetes due to a mutation of mitochondrial DNA that impairs the transfer of leucine into mitochondrial proteins has now been described in a large number of families, and results from impaired β cell function. The incidence of this disorder is as high as 1% to 3% in patients with diabetes in Japan and Korea, but lower in European populations. Most patients have a mild form of maternally transmitted diabetes with insulin deficiency that responds to oral hypoglycemic agents; however, some patients have a more severe clinical picture similar to type 1a diabetes. As many as 63% of patients with this subtype of diabetes have hearing loss and a smaller proportion (15%) have a syndrome of myopathy, encephalopathy, lactic acidosis, and stroke-like episodes.

Ketosis-Prone Diabetes

First described in young adult African American males from the Flatbush neighborhood in New York City, but since described in a number of populations of both African and non-African ancestry, these patients typically present with diabetic ketoacidosis and absolute insulin deficiency, followed by extended clinical remission off insulin. When tested during remission, however, maximal insulin secretory capacity remains markedly reduced, and these patients follow a relapsing course of DKA and hyperglycemia with eventual permanent insulin-deficient diabetes. These patients do not have islet cell autoantibodies or increased frequencies of HLA haplotypes associated with risk of autoimmune type 1a diabetes. Although there is often a family history of similar diabetes, the inheritance is not clearly Mendelian. Although KPD patients have been distinguished based on the history of relapsing β cell dysfunction, it remains controversial as to whether KPD represents a clinical entity distinct from other forms of nonautoimmune type 1b diabetes.

In a study of KPD patients of West African ancestry, linkage was established with coding variants in the *PAX4* gene. *PAX4* encodes for a transcription factor that functions downstream of *NEUROG3* in the formation of β cells. Fourteen to twenty-one percent of individuals of West African ancestry carry a coding variant that substitutes tryptophan for arginine at position 133 in *PAX4* and reduces its ability to repress transcription of target genes. Interestingly, the R133W variant is unique to people of West African ancestry. The R133W allele is approximately twice as frequent in individuals with KPD, and all individuals homozygous for R133W have KPD. An additional *PAX4* coding variant with reduced ability to bind to DNA was identified in one patient with KPD. Different coding variants in *PAX4* have also been identified in Thai families with MODY (MODY 9) and in Japanese

patients (both heterozygous and homozygous) with early onset insulin-deficient diabetes similar to KPD. Taken together, these data suggest that coding variants in the *PAX4* gene predispose to insulin-deficient diabetes, but the clinical phenotype may depend on the exact nature of the variant and interactions with other genes and environmental factors.

Genetic Defects of Insulin Action

These are rare and unusual causes of diabetes that result from mutations of the insulin receptor or from other genetically determined postreceptor abnormalities of insulin action. Metabolic abnormalities associated with these disorders may range from hyperinsulinemia and modest hyperglycemia to severe diabetes. Many individuals have acanthosis nigricans, polycystic ovaries with hyperandrogenism, and, in exceptional cases, pseudoacromegaly (discussed earlier).

Familial forms of insulin resistance associated with acanthosis nigricans have been termed type A insulin resistance, and are often associated with heterozygous mutations in the **insulin receptor** or homozygous mutations that retain some insulin-signaling capacity. Because of the capacity of the β cell to compensate for insulin resistance, these patients often do not present with hyperglycemia, but have severely elevated levels of circulating insulin.

If both copies of the insulin receptor gene carry mutations that nearly or completely abrogate signaling, the affected children present at birth with a syndrome known as **Leprechaunism** (or **Donohue syndrome**) that includes growth retardation, multiple developmental defects, lipoatrophy, severely elevated insulin levels, and hyperglycemia that does not respond to insulin therapy. These patients generally do not survive beyond a few weeks. Patients with **Rabson–Mendenhall syndrome** also have homozygous or compound heterozygous mutations in the insulin receptor, but with some small amount of residual signaling that results in a slightly less severe syndrome associated with abnormalities in the nails, teeth, and pineal gland. They may survive into adolescence.

Familial forms of type A insulin resistance without mutations in the insulin receptor have been described, and some of these may result from mutations in downstream components of the insulin-signaling cascade, although to date few such mutations have been found in humans. Variants in IRS-1 have been identified in patients with diabetes, but also occur in people with normal insulin sensitivity. Rare coding mutations in the genes encoding the p85α subunit of PI3 kinase, in AKT2, and in PPARγ have been associated with severe insulin resistance in a very small number of patients. The one patient with homozygous mutation of *AKT2* also had mild atrophy of subcutaneous limb adipose tissue.

Congenital and acquired forms of lipodystrophy can cause severe insulin resistance. Except in newborns with lipoatrophy secondary to complete loss of insulin receptors (Leprechaunism), alterations in the structure and function of the insulin receptor cannot be demonstrated in patients with insulin-resistant lipodystrophic diabetes, suggesting that the cause of the insulin resistance in these patients must reside in postreceptor pathways. Replacement of the adipokines leptin and adiponectin can reverse insulin resistance in mouse models of severe lipoatrophy, and leptin has been helpful in human cases of generalized lipodystrophy, demonstrating the importance of the adipocyte in regulating insulin function. Also, the loss of adipose storage depots in lipoatrophy leads to very high levels of circulating triglyceride-rich lipoprotein particles and increased deposition of fat in nonadipose tissues such as liver and muscle, which may contribute to dysfunction and insulin resistance in these tissues. The increase in ectopic fat deposition leads to profound hepatic steatosis in affected patients and can progress to cirrhosis and liver failure.

The congenital syndromes can be divided into generalized and partial lipodystrophies with or without dystrophic features in other tissues. The generalized syndromes are often identified in neonates, have severe insulin resistance at diagnosis and rapidly develop hyperglycemia, acanthosis nigricans, hyperandrogenism, and pseudoacromegaly. Recessive mutations causing generalized congenital lipodystrophy have been identified in three genes encoding proteins involved in the formation of lipid droplets (seipin and caveolin-1) and triglyceride (1-acylglycerol-3-phosphate-*O*-acyltransferase 2 [AGPAT2]).

Two syndromes of familial partial lipodystrophy without associated dysmorphic defects have been described in which the lipoatrophy usually first appears late in childhood, but may be proceeded by evidence of insulin resistance. The type 1 syndrome consists of atrophy of limb, gluteal and subcutaneous abdominal fat with sparing or increases of the abdominal visceral, upper trunk, head, and neck fat. In the type 2 syndrome, truncal and visceral fat are also affected and only vulval and head and neck depots are spared.

To date no genetic causes of the type 1 syndrome have been identified. In the type 2 syndrome, dominant mutations have been found in *LMNA*, the gene encoding the nuclear intermediate filament lamin A/C. In addition, one individual with the type 2 syndrome has been identified with a homozygous nonsense mutation in the gene encoding the lipid droplet protein CIDEC, and one individual was identified with a homozygous truncation of the gene encoding lipase maturation factor 1, a protein required for the maturation of both lipoprotein lipase and hepatic lipase.

Consistent with its known role in adipocytes differentiation, dominant mutations in the gene encoding PPARγ have also been described in patients with familial partial lipodystrophy and insulin resistance. These patients have a pattern of lipoatrophy similar to the type 2 syndrome with *LMNA* mutations, but with less severe subcutaneous fat loss, and have been labeled type 3 familial partial lipodystrophy. Less commonly, mutations in the PLIN1, AKT2 and CIDEC genes have been associated with familial forms of partial lipodystrophy.

Several syndromes of lipodystrophy associated with other dysmorphic features have been described. Autosomal recessive mutations in *ZMPSTE24*, which encodes a metalloproteinase that cleaves prolamin A to produce mature lamin A, cause generalized lipodystrophy associated with mandibuloacral dysplasia. In patients with partial lipodystrophy combined with mandibuloacral dysplasia, autosomal recessive mutations have been identified in *LMNA*, but these mutations are distinct from those that cause familial partial lipodystrophy type 2. Remarkably, yet a different set of mutations in *LMNA* cause the autosomal recessive syndrome Hutchinson–Gilford progeria, which includes severe early

onset generalized lipodystrophy. Additional distinct mutations in *LMNA* cause several other congenital dysmorphic syndromes, including muscular dystrophies, familial dilated cardiomyopathy, and Charcot-Marie-Tooth disease.

Acquired forms of partial and generalized lipodystrophy with insulin resistance and diabetes can develop secondary to infections, autoimmunity, paraneoplastic syndromes, collagen vascular disorders, drugs, or unknown causes. One common form is seen in patients with HIV infection following treatment with protease inhibitors.

Neonatal Diabetes

Neonatal diabetes, defined as diabetes diagnosed before 6 months of age, is rare, occurring in fewer than 1 in 200,000 live births. Children with neonatal diabetes often present with decreased birth weight (intrauterine growth restriction, IUGR) and decreased fat stores in addition to hyperglycemia. Most commonly these children have reduced circulating insulin and C-peptide levels caused by inherited β cells defects, although rare inherited defects in insulin signaling can also present in neonates. In approximately half of cases, neonatal diabetes is transient: the diabetes goes into remission (normoglycemia with no therapy) before 18 months of age, although it usually returns at puberty. In the remainder of cases, the diabetes is permanent.

Transient neonatal diabetes (TNDM) Defects in imprinting underlie most cases of TNDM. The most common genetic defect is paternal uniparental isodysomy (replacement of the maternal copy of the region with the paternal copy) of an imprinted region at chromosome 6q24. In this region, the maternal copy of the chromosome is normally silenced by methylation. Simple duplication of the paternal copy of this region, or mutations in ZFP57, a zinc-finger protein required globally for the methylation of imprinted regions, can give the same TNDM phenotype. The affected region of 6q24 contains two genes of uncertain function, but one of them, *ZAC*, has been identified as a tumor suppressor gene and may activate the expression of an inhibitor of cell proliferation. An increase in hypomethylated copies of *ZAC* may lead to the inhibition of β cell proliferation and inadequate β cell mass.

Some autosomal dominant activating mutations in *KCNJ11* and *ABCC8*, the genes encoding the two subunits of the ATP-sensitive potassium channels in β cells (see Figure 17–5) and listed among autosomal dominant forms of heritable diabetes (discussed earlier) can cause TNDM. Autosomal dominant mutations in the MODY 5 gene, *HNF1*β, can also cause TNDM.

Permanent neonatal diabetes (PNDM) Autosomal dominant activating mutations in *KCNJ11* and *ABCC8* can also cause PNDM. Some of the autosomal dominant insulin mutations that cause rapid β cell apoptosis will result in permanent diabetes in neonates. In addition, most of the syndromes caused by autosomal recessive genetic defects in β cells (see Table 17–9) present in neonates. These include mutations in the MODY genes *GCK*, *PDX1*, and *NEUROD1*, the transcription factor genes *NKX2-2*, *MNX1*, *PTF1A*, *GLIS3*, and *RFX6*, and the Wolcott–Rallison

syndrome gene *EIF2AK3*. The IPEX syndrome (see Genetics of Type 1 Diabetes, discussed earlier), caused by mutations in *FOXP3*, can also present with accelerated autoimmune type 1 diabetes in neonates. Finally, mutations that cause complete, or nearly complete, loss of insulin signaling (Leprechaunism) also present with PNDM.

Several other syndromes of neonatal diabetes of unknown etiology associated with a variety of other developmental defects have been described. With the advent of rapid whole genome sequencing, the genetic defects that cause these syndromes, as well as later onset diabetes, may soon be identified, providing further insights into the pathogenesis of diabetes and identifying potential therapeutic targets.

Monogenic Autoimmune Syndromes

Several syndromes of altered immune function with multiple endocrine gland involvement caused by single gene mutations have been described. The APS1 and IPEX syndromes are described in the section on the genetics of type 1 diabetes and in Chapter 2. Recently several additional monogenic autoimmune syndromes that include type 1 diabetes have been identified with autosomal dominant (*STAT1*, *STAT3*, and *SIRT1*) or recessive (*IL2RA*, *ITCH*, and *LRBA*) inheritance.

Other Genetic Syndromes Sometimes Associated with Diabetes

More than 50 distinct genetic syndromes involve an increased incidence of diabetes mellitus. These include the chromosomal abnormalities of Down syndrome, Klinefelter syndrome, and Turner syndrome. In addition, a number of complex syndromes associated with neuromuscular pathologies (Freidriech ataxia, Huntington chorea, porphyria, muscular dystrophies) or severe obesity (Laurence–Moon–Biedl, Bardet–Biedl, and Prader–Willi syndromes) have been associated with diabetes.

SECONDARY DIABETES

Diabetes due to Diseases of the Exocrine Pancreas

Any process that diffusely damages or substantially displaces the pancreas can cause diabetes, although individuals with a predisposition to type 2 diabetes are probably more susceptible to developing diabetes with lesser degrees of pancreatic involvement. Because glucagon-secreting α cells are also damaged or removed by these processes, less insulin is usually required for replacement—as compared with most other forms of diabetes that leave α cells intact.

Acquired causes include pancreatitis, trauma, infection, pancreatic carcinoma, and pancreatectomy. Fibrocalculous pancreatopathy, a form of acquired pancreatitis with extensive fibrosis and ductal calculi seen commonly in tropical regions, may result from both dietary and genetic contributors, although the exact cause remains obscure. Like chronic pancreatitis from a variety of causes, fibrocalculous involvement of the pancreas may be accompanied

by abdominal pain radiating to the back and associated with pancreatic calcifications on x-ray. When extensive enough, hemochromatosis and cystic fibrosis can also displace β cells and reduce insulin secretion. Autosomal dominant mutations in carboxyl-ester lipase (CEL), an exocrine enzyme, cause accelerated exocrine pancreatic damage and diabetes at a young age, and have been designated as MODY 8 (see Table 17–8).

Endocrinopathies

Excess production of certain hormones—GH (acromegaly), glucocorticoids (Cushing syndrome or disease), catecholamines (pheochromocytoma), thyroid hormone (thyrotoxicosis), glucagon (glucagonoma), or pancreatic somatostatin (somatostatinoma)—can produce relative insulin deficiency and diabetes by a number of mechanisms. In all but the last instance (somatostatinoma), peripheral responsiveness to insulin is impaired. In addition, excess of catecholamines or somatostatin decreases insulin release from β cells. Diabetes mainly occurs in individuals with underlying defects in insulin secretion, and hyperglycemia typically resolves when the hormone excess is corrected.

Drug- or Chemical-Induced Diabetes

Many drugs are associated with carbohydrate intolerance or frank diabetes mellitus. Some act by interfering with insulin release from the β cells (thiazides, phenytoin, cyclosporine), some by inducing insulin resistance (glucocorticoids, oral contraceptive pills, niacin, and antiviral protease inhibitors), and some by causing β cell destruction (intravenous pentamidine). Atypical antipsychotic medications can provoke substantial weight gain and insulin resistance, but the high reported incidence of diabetic ketoacidosis in patients on these drugs suggests that they may also impair β cell function.

The increasing therapeutic use of drugs that block inhibitory immune system pathways that keep immune responses in check has led to the development of new syndromes of autoimmunity including autoimmune diabetes (essentially a secondary form of type 1 diabetes). Both α interferon used to treat Hepatitis C and immune checkpoint inhibitors such the monoclonal antibodies against PD-1 and CTLA-4 used to treat cancer have been associated with development of a fulminant form of autoimmune diabetes similar to type 1 diabetes, as well as other endocrine and non-endocrine autoimmune disorders.

Adrenergic drugs impact glucose metabolism in complex and often opposing ways because of differing effects on insulin secretion, glucagon secretion, hepatic glucose output, peripheral insulin sensitivity, and weight gain. In clinical practice, the first generation, nonselective β blockers such as propanolol tend to modestly increase glucose levels, at least in part due to increases in insulin resistance, but potentially by decreasing insulin secretion as well. The second generation, selective β1 blockers also tend to increase blood glucose, but the third generation drugs with combined α and β blockade have minimal effects on blood glucose. In contrast, nonselective α agonist and α2 agonists tend to raise blood glucose, probably due to their combined effects on insulin

secretion and hepatic glucose output. However, although β blockers and α2 agonists like clonidine, as well as calcium channel blockers, inhibit glucose-induced insulin release from *in vitro* preparations of pancreatic β cells, these drugs have minimal or modest effects on blood glucose control at the levels used in standard antihypertensive therapy in humans.

Finally it must be kept in mind that the most common toxin causing diabetes is ethanol. Chronic alcoholic pancreatitis with secondary loss of β cells accounts for approximately 1% of diabetes in the United States.

Infections Causing Diabetes

Certain viruses have been associated with direct pancreatic β cell destruction in animals. Diabetes is also known to develop frequently in humans who had congenital rubella, although most of these patients have HLA and immune markers characteristic of type 1 diabetes. In addition, coxsackievirus B, cytomegalovirus, adenovirus, and mumps have been implicated in inducing certain cases of diabetes.

Uncommon Forms of Immune-Mediated Diabetes

A severe form of insulin resistance has been reported in patients who developed high titers of antibodies that bind to the insulin receptor and block the action of insulin in its target tissues. As in other states of extreme insulin resistance, these patients often have acanthosis nigricans. In the past, this form of immune-mediated diabetes was termed type B insulin resistance. Most commonly these antibodies are of idiopathic origin, but they have also been described in monoclonal gamopathies and multiple myeloma and in patients with ataxia telangectasia syndrome.

Patients with POEMS, a syndrome of plasma cell dyscrasia associated with polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes, have an increased incidence of diabetes as well as other endocrine disorders. The cause of the diabetes in these patients has not been established.

CLINICAL FEATURES OF DIABETES MELLITUS

The principal clinical features of the two major types of diabetes mellitus are listed for comparison in Table 17–10.

Type 1 Diabetes

Patients with type 1 diabetes present with symptoms and signs related to hyperglycemia and hyperketonemia. The severity of the insulin deficiency and the acuteness with which the catabolic state develops determine the intensity of the osmotic and ketotic excess.

A. Symptoms   Increased urination is a consequence of osmotic diuresis secondary to sustained hyperglycemia. This results in a loss of glucose as well as free water and electrolytes in the urine. Nocturnal enuresis due to polyuria may signal the onset of

TABLE 17–10 Clinical features of diabetes at diagnosis.

	Diabetes Type 1	Diabetes Type 2
Polyuria and thirst	++	+
Weakness or fatigue	++	+
Polyphagia with weight loss	++	−
Recurrent blurred vision	+	++
Vulvovaginitis or pruritus	+	++
Peripheral neuropathy	+	++
Nocturnal enuresis	++	−
Often asymptomatic	−	++

diabetes in very young children. Thirst is a consequence of the hyperosmolar state, as is blurred vision, which often develops as the lenses and retinas are exposed to hyperosmolar fluids.

Weight loss, despite normal or increased appetite, is a common feature of type 1 diabetes when it develops subacutely over a period of weeks. The weight loss is initially due to depletion of water, glycogen, and triglyceride stores. Chronic weight loss due to reduced muscle mass occurs as amino acids are diverted to form glucose and ketone bodies.

Lowered plasma volume produces dizziness and weakness due to postural hypotension when sitting or standing. Total body potassium loss and the general catabolism of muscle protein contribute to the weakness.

Paresthesias may be present at the time of diagnosis of type 1 diabetes, particularly when the onset is subacute. They reflect a temporary dysfunction of peripheral sensory nerves and usually clear, as insulin replacement restores glycemic levels closer to normal; thus, their presence suggests neurotoxicity from sustained hyperglycemia.

When insulin deficiency is severe and of acute onset, the above symptoms progress in an accelerated manner. Ketoacidosis exacerbates the dehydration and hyperosmolality by producing anorexia, nausea, and vomiting, thus interfering with oral fluid replacement. As plasma osmolality exceeds 330 mOsm/kg (normal, 285-295 mOsm/kg), impaired consciousness ensues. With progression of acidosis to a pH of 7.1 or less, deep breathing with a rapid ventilatory rate (Kussmaul respiration) occurs as the body attempts to eliminate carbonic acid. With worsening acidosis (to pH 7.0 or less), the cardiovascular system may be unable to maintain compensatory vasoconstriction; severe circulatory collapse may result.

B. Signs   The patient's level of consciousness can vary depending on the degree of hyperosmolality. When insulin deficiency develops relatively slowly and sufficient water intake is maintained to permit renal excretion of glucose and appropriate dilution of extracellular sodium chloride concentration, patients remain relatively alert and physical findings may be minimal. When vomiting occurs in response to worsening ketoacidosis, dehydration progresses and compensatory mechanisms become inadequate to keep plasma osmolality below 330 mOsm/kg. Under these circumstances, stupor or even coma may occur. Evidence of dehydration

in a stuporous patient, with rapid deep breathing and the fruity breath odor of acetone, suggests the diagnosis of diabetic ketoacidosis.

Postural hypotension indicates a depleted plasma volume; hypotension in the recumbent position is a serious prognostic sign. Loss of subcutaneous fat and muscle wasting are features of more slowly developing insulin deficiency. In occasional patients with slow, insidious onset of insulin deficiency, subcutaneous fat may be considerably depleted. An enlarged liver, eruptive xanthomas on the flexor surface of the limbs and on the buttocks, and lipemia retinalis indicate that chronic insulin deficiency has resulted in chylomicronemia, with elevated circulating triglycerides, usually to over 2000 mg/dL (Chapter 19).

Type 2 Diabetes

Patients with type 2 diabetes usually have less severe insulin deficiency than type 1 patients and the symptoms and signs at presentation reflect this difference.

A. Symptoms   Many patients with type 2 diabetes have an insidious onset of hyperglycemia and may be relatively asymptomatic initially. The diagnosis may be made only after glycosuria or hyperglycemia is noted during routine laboratory studies. Chronic skin infections are common. Generalized pruritus and symptoms of candidal vaginitis are frequently the initial complaints of women with type 2 diabetes. Men may complain of an itchy rash of the prepuce. Some patients can remain undiagnosed for many years and the initial presentation may be due to complications such as visual disturbance due to retinopathy or foot pain or infection due to a peripheral neuropathy. Patients with a more severe insulin deficiency have the classical symptoms of polyuria, thirst, blurred vision, paresthesias, and fatigue. This is especially true in individuals who consume large amounts of carbohydrate-rich fluids in response to the thirst.

B. Signs   Many patients are obese or overweight. Even those patients who are not significantly overweight often have a characteristic fat distribution with more fat in the upper part of the body (particularly the abdomen, chest, neck, and face) and relatively less fat on the appendages, which may be quite muscular (the *metabolically obese*). This centripetal fat distribution has been termed **android** and is characterized by a high waist circumference. It differs from the more centrifugal **gynecoid** form of obesity, in which fat is localized more in the hips and thighs and less in the upper parts of the trunk. A larger waist circumference increases the risk for diabetes for any given BMI. Thus in patients with the metabolic syndrome, a waist circumference more than 40 in (102 cm) in men and more than 35 in (88 cm) in women is associated with an increased risk of diabetes. MRI and CT scans reveal that these patients with increased waist circumference have accumulation of fat in the omental and mesenteric distributions. This visceral fat correlates with insulin resistance, whereas fat predominantly in subcutaneous tissues of the abdomen has little, if any, association with insulin insensitivity.

Some patients, especially the obese, may have acanthosis nigricans—hyperpigmented, hyperkeratotic skin in the axilla, groin, and back of neck. This sign is associated with significant insulin resistance. Hypertension may be present especially in the obese patient. Eruptive xanthomata on the flexor surface of the limbs and on the buttocks and lipemia retinalis due to hyperchylomicronemia can occur in patients with uncontrolled type 2 diabetes who also have a familial form of hypertriglyceridemia. In women, candidal vaginitis with a reddened, inflamed vulvar area and a profuse whitish discharge may herald the presence of diabetes. In men, candidal infection of the penis may lead to reddish appearance of the penis and/or prepuce with eroded white papules and a white discharge. The occasional patient who has had undiagnosed diabetes for some time may present with retinopathy or peripheral neuropathy.

Patients can also present in hyperglycemic hyperosmolar coma—profoundly dehydrated, hypotensive, lethargic, or comatose without Kussmaul respirations.

Laboratory Testing in Diabetes Mellitus

Tests of urine glucose and ketone bodies, as well as whole blood or plasma glucose measured in samples obtained under basal conditions and after glucose administration, are very important in evaluation of the patient with diabetes. Tests for glycosylated hemoglobin have proved useful in both initial evaluation and in assessment of the effectiveness of therapeutic management. In certain circumstances, measurements of insulin or C-peptide levels and levels of other hormones involved in carbohydrate homeostasis (eg, glucagon, GH) may be useful. In view of the increased risk of atherosclerosis in patients with diabetes, determination of serum total cholesterol, HDL-cholesterol, triglycerides, and LDL-cholesterol may be helpful.

Urine Glucose

Several problems are associated with using urine glucose as an index of blood glucose, regardless of the method employed. First of all, the glucose concentration in bladder urine reflects the blood glucose at the time the urine was formed. Therefore, the first voided specimen in the morning contains glucose that was excreted throughout the night and does not reflect the morning blood glucose at all. Some improvement in the correlation of urine glucose to blood glucose can be obtained if the patient *double voids*—that is, empties the bladder completely, discards that sample, and then urinates again about one-half hour later, testing only the second specimen for glucose content. However, difficulty in completely emptying the bladder (large residual volumes), problems in understanding the instructions, and the inconvenience impair the usefulness of this test. Self-monitoring of blood glucose has replaced urine glucose testing in most patients with diabetes (particularly those receiving insulin therapy).

Several commercial products are available for determining the presence and amount of glucose in urine. The older and more cumbersome bedside assessment of glycosuria with Clinitest tablets has generally been replaced by the dipstick method, which is rapid, convenient, and glucose specific. This method consists of paper strips (Clinistix, Tes-Tape) impregnated with enzymes (glucose oxidase and hydrogen peroxidase) and a chromogenic dye that is colorless in the reduced state. Enzymatic generation of hydrogen peroxide oxidizes the dye to produce colors whose intensity depends on the glucose concentration. These dipsticks are sensitive to as little as 0.1% glucose (100 mg/dL) but do not react with the smaller amounts of glucose normally present in nondiabetic urine. The strips are subject to deterioration if exposed to air, moisture, and extreme heat and must be kept in tightly closed containers except when in use. False-negative results may be obtained in the presence of alkaptonuria and when certain substances such as salicylic acid or ascorbic acid are ingested in excess. All these false-negative results occur because of the interference of strong reducing agents with oxidation of the chromogen.

Although glycosuria reflects hyperglycemia in more than 90% of patients, two major classes of nondiabetic glycosuria must be considered:

A. Alterations in renal handing of glucose Disorders associated with abnormalities in renal glucose handling include Fanconi syndrome, a group of disorders characterized by combined renal wasting of multiple solutes including amino acids, uric acid, phosphate, and bicarbonate as well as glucose and caused by both genetic and acquired defects of the proximal renal tubule. Familial renal glycosuria, a benign inherited disorder manifest only by persistent glycosuria in the setting of euglycemia is caused by mutations in SGLT2, the sodium-glucose cotransporter responsible for the bulk of glucose reabsorption in the proximal tubule.

In addition, glycosuria is relatively common in pregnancy as a consequence of the increased load of glucose presented to the tubules by the elevated glomerular filtration rate during pregnancy. As many as 50% of pregnant women normally have demonstrable sugar in the urine, especially after the first trimester.

B. Excretion of sugars other than glucose in the urine Occasionally, a sugar other than glucose is excreted in the urine. Lactosuria during the late stages of pregnancy and the period of lactation is the most common example. Much rarer are other conditions in which inborn errors of metabolism allow fructose, galactose, or a pentose (1-xylose) to be excreted in the urine. Testing the urine with glucose-specific strips helps differentiate true glucosuria from other glycosurias.

Microalbuminuria and Proteinuria

Urinary albumin can now be detected in microgram concentrations using high-performance liquid chromatography or immunoassay methodology that is more sensitive than previous available tests. Conventional 24-hour urine collections, in addition to being inconvenient for patients, also show wide variability of albumin excretion, because several factors such as sustained upright posture, dietary protein, and exercise tend to increase albumin excretion rates. For these reasons, it is preferable to measure the albumin-creatinine ratio in an early morning spot urine collected

on awakening—prior to breakfast or exercise—and brought in by the patient for laboratory analysis. A ratio of albumin (μg/L) to creatinine (mg/L) of less than 30 is normal, and a ratio of 30 to 300 indicates abnormal microalbuminuria. Values greater than 300 are referred to as macroalbuminuria.

The minimal detection limit of protein on a standard urine dipstick is 10 to 20 mg/dL. If the dipstick is positive then it is likely that the patient has microalbuminuria and this should be specifically tested. The information from a spot urine sample is adequate for diagnosis and treatment and it is not usually necessary to perform a 24-hour urine collection for protein loss and creatinine clearance.

Blood Glucose Testing

A. Venous blood samples in the laboratory
Venous glucose samples should be collected in tubes containing sodium fluoride, which inhibits enolase and prevents glycolysis in the blood sample that would artifactually lower the measured glucose level. In the absence of fluoride, the rate of disappearance of glucose in the presence of blood cells has been reported to average 10 mg/dL/h—the rate increases with glucose concentration, temperature, and white blood cell count. Fluoride takes about 1 hour to effectively stop glycolysis. Therefore, the rate of decline during the first hour is the same in tubes with or without fluoride. A very high white blood cell count will lower glucose levels even in the presence of fluoride. Ideally the blood should be collected in a sodium fluoride/potassium oxalate tube, placed on ice and the plasma separated from the cells within 60 minutes.

Plasma or serum from venous blood samples has the advantage over whole blood of providing values for glucose that are independent of hematocrit and reflect levels in the interstitial spaces to which body tissues are exposed. For these reasons—and because plasma and serum lend themselves to automated analytic procedures—they are used in most laboratories. The glucose concentration is 10% to 15% higher in plasma or serum than in whole blood because structural components of blood cells are absent. Whole blood glucose determinations are seldom used in clinical laboratories, but are used by patients using home blood glucose monitors.

Glucose levels can be measured in the laboratory using enzymatic methods (such as glucose oxidase or hexokinase), condensation methods (such as *o*-toluidine), or reducing methods. The reducing methods take advantage of the reducing properties of glucose to change the redox state of a metal ion; however, the method is nonspecific and any strong reducing agent can cross-react to yield spuriously elevated glucose values. In condensation methods, the aldehyde group of glucose undergoes condensation with aromatic compounds to yield a colored product. In the most commonly used condensation reaction, *o*-toluidine reacts with glucose to form a glucosamine that has an intense green color. The color is measured spectrophotometrically to estimate the glucose concentration. *o*-Toluidine, however, has the drawback of being highly corrosive and toxic. In the enzymatic method, glucose oxidase reacts with glucose, water, and oxygen to form gluconic acid and hydrogen peroxide. The hydrogen peroxide can then be used

TABLE 17–11 Criteria for the diagnosis of diabetes.

	Normal Glucose Tolerance	Prediabetes	Diabetes Mellitus[a]
Fasting plasma glucose mg/dL(mmol/L)[b]	<100 (5.6)	100-125 (5.6-6.9) (impaired fasting glucose)	≥126 (7.0)
2 h after glucose load mg/dL (mmol/L)[c]	<140 (7.8)	≥140-199 (7.8-11.0) (impaired glucose tolerance)	≥200 (11.1)
HbA$_{1c}$ (%)[a]	<5.7	5.7-6.4	≥6.5
Symptoms and random glucose level (mg/dL)	–	–	≥200

[a]A fasting plasma glucose or HbA$_{1c}$ is diagnostic of diabetes if confirmed by repeat testing. HbA$_{1c}$ test should be performed using an assay certified by the National Glycohemoglobin Standardization program and standardized to the DCCT assay.

[b]A fasting plasma glucose ≥126 mg/dL is diagnostic of diabetes if confirmed on a subsequent day to be in the diabetic range after an overnight fast.

[c]Give 75 g of glucose dissolved in 300 mL of water after an overnight fast in subjects who have been receiving at least 150 to 200 g of carbohydrate daily for 3 days before the test. In the absence of unequivocal hyperglycemia, the result should be confirmed by repeat testing.

to oxidize a chromogen or the consumption of oxygen measured to estimate the amount of glucose present. Current laboratories use enzymatic methods to determine glucose levels.

The range of normal fasting plasma or serum glucose is 70 to 100 mg/dL (3.9-5.5 mmol/L). A plasma glucose level of 126 mg/dL (7.0 mmol/L) or higher on more than one occasion after at least 8 hours of fasting is diagnostic of diabetes mellitus (Table 17–11). Fasting plasma glucose levels of 100 mg/dL (5.6 mmol/L) to 125 mg/dL (6.9 mmol/L) are associated with increased risk for diabetes (impaired fasting glucose).

B. Self monitoring of blood glucose
Capillary blood glucose measurements performed by patients themselves, as outpatients, are extremely useful. In type 1 patients in whom tight metabolic control is attempted, they are indispensable. There are several paper strip (glucose oxidase; hexokinase; glucose dehydrogenase with nicotinamide adenine dinucleotide, glucose dehydrogenase with flavin-adenine dinucleotide, glucose dehydrogenase with pyrroloquinoline quinine) methods for measuring glucose on capillary blood samples. A reflectance photometer or an amperometric system is then used to measure the reaction that takes place on the reagent strip. A large number of blood glucose meters are now available. All are accurate, but they vary with regard to speed, convenience, size of blood samples required, and cost. Popular models include those manufactured by Life-Scan (One Touch), Bayer Corporation (Breeze, Contour), Roche Diagnostics (Accu-Chek), and Abbott Laboratories (Precision, FreeStyle). Various glucometers appeal to a particular consumer need and are relatively inexpensive, ranging from $50.00 to $100.00 (U.S.) each. Test strips remain a major expense, costing $0.25 to $1.50 (U.S.) each.

The meters have memories and can compute blood glucose averages. The data can be downloaded into a computer. Some meters are designed to communicate with a specific insulin pump. Contour Next Link meter, for example, communicates with the MiniMed Medtronic pump. In self-monitoring of blood glucose, patients must prick a finger with 26- to 33-gauge lancets. This can be facilitated by a small plastic trigger device such as an Accu-Chek multiclix (Roche Diagnostics), Microlet (Boehringer-Mannheim), or one touch lancing device (Lifescan, Inc.).

The clinician should be aware of the limitations of the self-glucose monitoring systems. The strips have limited lifespans and improper storage (high temperature; open vial) can affect their function. Patients should also be advised not to use expired strips. Some of the older meters require input of a code for each batch of strips and failure to enter the code can result in misleading results. Most meters no longer require this step. Increases or decreases in hematocrit can decrease or increase the measured glucose values, respectively. The mechanism underlying this effect is not known, but presumably it is due to the impact of red cells on the diffusion of plasma into the reagent layer. Meters and the test strips are calibrated over glucose concentrations ranging from 60 to 160 mg/dL, and the accuracy is not as good for higher and lower glucose levels. Thus, when the glucose is less than 60 mg/dL, the difference between the meter and the laboratory value may be as much as 20%. Glucose-oxidase-based amperometric systems underestimate glucose levels in the presence of high oxygen tension. This may be important in critically ill patients who are on supplemental oxygen, and under these circumstances, a glucose dehydrogenase-based system may be preferable. Glucose-dehydrogenase pyrroloquinoline quinone (GDH-PQQ) systems may report falsely high glucose levels in patients who are receiving parenteral products containing nonglucose sugars such as maltose, galactose, or xylose or their metabolites. Some meters have been approved for measuring glucose in blood samples obtained at alternative sites such as the forearm and thigh. There is, however, a 5- to 20-minute lag in the glucose response on the arm with respect to the glucose response on the finger. Forearm blood glucose measurements could therefore result in a delay in detection of rapidly developing hypoglycemia. Impaired circulation to the fingers (for example, in patients with Raynaud's disease) will artificially lower finger-stick glucose measurements (pseudohypoglycemia). The accuracy of data obtained by glucose monitoring requires education of the patient in sampling and measuring procedures as well as in proper calibration of the instruments. Bedside glucose monitoring in a hospital setting requires rigorous quality control programs and certification of personnel to avoid errors.

Continuous Glucose Monitoring Systems

A number of continuous glucose monitoring (CGM) systems are currently available for clinical use. The systems manufactured by Medtronic MiniMed, DexCom, Abbott Diagnostics involve inserting a subcutaneous biosensor (rather like an insulin pump cannula) that measures glucose concentrations in the interstitial fluid for 3 to 14 days. The glucose values are available for review by the patient at time of measurement. The systems differ in how the glucose data are displayed. The MiniMed system can only be used in conjunction with the MiniMed pump and the glucose data are displayed on the screen of the pump. The DexCom system wirelessly transmits glucose data to a separate pager-like device with a screen. The DexCom system also has the option of having the data appear on smart phones or smart watches or on the screens of several insulin pumps. The FreeStyle Libre (Abbott Diagnostics) sensor system requires the patient to hold a reading device close to the sensor patch for about a second to see the real time glucose value. The MiniMed and DexCom systems also display directional arrows indicating rate and direction of change and alarms can be set for dangerously low or high glucose values. Patients still have to calibrate the MiniMed and DexCom devices with periodic fingerstick glucose levels, and since there are concerns regarding reliability, it is still necessary to confirm the displayed glucose level with a fingerstick glucose before making interventions such as injecting extra insulin or eating extra carbohydrates. The FreeStyle Libre system does not require calibration and glucose readings are available 1 hour after insertion of sensors. The glucose data can be downloaded into computers for review by patients and their medical providers.

A 6-month randomized controlled study of type 1 patients showed that adults (25 years and older) using these systems had improved glycemic control without an increase in the incidence of hypoglycemia. A randomized controlled study of CGM during pregnancy showed improved glycemic control in the third trimester, lower birth weight, and reduced risk of macrosomia. The individual glucose values are not that critical—what matters is the direction and the rate at which the glucose is changing, allowing the user to take corrective action. The wearer also gains insight into the way particular foods and activities affect their glucose levels. The other main benefit is the low glucose alert warning. The MiniMed insulin pump can be programmed to automatically suspend insulin delivery for up to 2 hours when the glucose levels on its CGM device falls to a preset level and the patient does not respond to the alert. This insulin suspension feature has been shown to reduce the amount of time patients are in the hypoglycemic range at night. Many of these systems are covered by insurance. The initial cost is about $800 to $1000, and the sensor, which has to be changed every 3 to 14 days, costs $35 to $77; the out-of-pocket expense is about $4000 annually.

There is great interest in using the data obtained from these CGM systems to automatically deliver insulin by continuous subcutaneous insulin infusion pump. Algorithms have been devised to link CGM to insulin delivery. These closed loop systems (artificial pancreas) have been shown in short-term clinical studies to improve night time glucose control and reduce the risk of nocturnal hypoglycemia. The first of these closed loop systems (MiniMed system) has been approved for clinical use. With this system, the continuous glucose monitor readings are used to automatically adjust basal insulin rates by the insulin pump.

Urine and Serum Ketone Determinations

In the absence of adequate insulin, three major ketone bodies are formed and excreted into the urine: β-hydroxybutyrate (often the

most prevalent in diabetic ketoacidosis), acetoacetate, and acetone. Acetone and acetoacetate react with sodium nitroprusside (nitroferricyanide) in the presence of alkali to produce a purple-colored complex. Ketone test strips (ketostix; Keto-Diastix) utilize this nitroprusside reaction to quantify acetone and acetoacetate levels in urine and plasma. A strongly positive reaction in undiluted serum correlates with a serum ketone concentration of at least 4 mmol/L. Although these tests do not detect β-hydroxybutyric acid, which lacks a ketone group, the semiquantitative estimation of the other ketone bodies is nonetheless usually adequate for clinical assessment of ketonuria. Ketostix and Keto-Diastix have short shelf-lives (90 days) once the containers are opened and using expired strips can give false-negative results. It is better therefore to buy individually foil wrapped strips.

Many laboratories measure β-hydroxybutyrate levels and there are meters available (Precision Xtra, Nova Max Plus) for patients to measure β-hydroxybutyrate levels on capillary glucose samples. This technology uses hydroxybutyrate dehydrogenase to catalyse the oxidation of β-hydroxybutyrate to acetoacetate with concomitant reduction of NAD^+ to NADH. The NADH is reoxidized to NAD^+ by a redox mediator and a current is generated that is directly proportional to β-hydroxybutyrate concentration. β-Hydroxybutyrate levels more than 0.6 nmol/L require evaluation. Levels more than 3.0 nmol/L, which is equivalent to very large urine ketones, will require hospitalization.

Other conditions besides diabetic ketoacidosis may cause ketone bodies to appear in the urine; these include starvation, high-fat diets, alcoholic ketoacidosis, fever, and other conditions in which metabolic requirements are increased.

Glycated Hemoglobin Assays

Ketoamine reactions between glucose and other sugars and free amino groups on the alpha and beta chain lead to glycated forms of hemoglobin. Only glycation of the *N*-terminal valine of the beta chain imparts sufficient negative charge to the hemoglobin molecule to allow separation by charge-dependent techniques. The charge-separated hemoglobins are collectively referred to as hemoglobin A_1 (HbA_1). The major form of HbA_1 is hemoglobin A_{1c} (HbA_{1c}), where glucose is the carbohydrate. This form comprises 4% to 6% of total hemoglobin A. The remaining HbA_1 species contain fructose 1,6-diphosphate (HbA_{1a1}), glucose 6-phosphate (HbA_{1a2}), and an unknown carbohydrate moiety (HbA_{1b}). The hemoglobin A_{1c} fraction is abnormally elevated in diabetic patients with chronic hyperglycemia. Some laboratories measure the sum of these glycohemoglobins (GHbs) and report the total as hemoglobin A_1, but most laboratories have converted to the highly specific HbA_{1c} assay. Methods for measuring HbA_{1c} include electrophoresis, cation exchange chromatography, boronate affinity chromatography, and immunoassays Office-based immunoassays using capillary blood give a result in about 9 minutes, and this allows for immediate feedback to the patients regarding their glycemic control.

The red cell lifespan is about 120 days and so HbA_{1c} levels reflect glucose levels in the preceding 8 to 12 weeks. The value, however, is weighted to more recent glucose levels (previous month)

and this explains why significant changes in HbA_{1c} are observed with short-term (1 month) changes in mean plasma glucose levels. Measurements should be made in patients with either type of diabetes mellitus at 3- to 4-month intervals so that adjustments in therapy can be made if GHb is either subnormal or if it is more than 1% above the upper limits of normal for a particular laboratory. In patients monitoring their own blood glucose levels, GHb values provide a valuable check on the accuracy of monitoring. In patients who do not monitor their own blood glucose levels, GHb values are essential for adjusting therapy.

The various HbA_{1c} assays have been standardized to the assay used in the Diabetes Control and Complications Trial (DCCT) allowing the results to be related to the risks of developing microvascular complications. There is a linear relationship between the HbA_{1c} value and average glucose. The A_{1c}-Derived Average Glucose Study collected 3 months of blood glucose data on 507 subjects—normals, type 1 and type 2 diabetics. The estimated average glucose was calculated by combining weighted results from 2 days of CGM per month and seven point capillary blood glucose profiles (preprandial, postprandial, and bedtime) for at least 3 days/wk. The HbA_{1c} was measured at the end of the 3 months. The relationship between average glucose and HbA_{1c} based on linear regression analysis was Average glucose = (28.7 × HbA_{1c}) – 46.7. There is, however, substantial intra-individual variability. For HbA_{1c} values between 6.9% and 7.1%, the glucose levels ranged from 125 to 205 mg/dL (6.9-11.4 mmol/L; 95% CIs). For HbA_{1c} of 6%, the mean glucose levels ranged from 100 to 152 mg/dL (5.5-8.5 mmol/L); and for 8% they ranged from 147 to 217 mg/dL (8.1-12.1 mmol/L).

The accuracy of HbA_{1c} values can be affected by hemoglobin variants or derivatives, the effect depending on the specific hemoglobin variant or derivative and the specific assay used. Immunoassays that use an antibody to the glycated amino terminus of β globin do not recognize the terminus of the γ globin of hemoglobin F, and so in patients with high levels of hemoglobin F, immunoassays give falsely low estimates of HbA_{1c}. Cation exchange chromatography separates hemoglobin species by charge differences. Therefore, hemoglobin variants that coelute with HbA_{1c} can lead to an overestimation of the HbA_{1c} value. Chemically modified derivatives of hemoglobin such as carbamylated (in renal failure) or acetylated (high-dose aspirin therapy) hemoglobin can, in some methods, coelute with HbA_{1c}. Since many hemoglobin variants do not have a clinical phenotype, they should be considered if there is discordance between home blood glucose and HbA_{1c} measurements. The National Glycohemoglobin Standardization Program website (www.ngsp.org) has information on the impact of frequently encountered hemoglobin variants and traits on the results obtained with the commonly used HbA_{1c} assays.

Any condition that shortens erythrocyte survival or decreases mean erythrocyte age (eg, recovery from acute blood loss, hemolytic anemia) falsely lower HbA_{1c} irrespective of the assay method used. Alternative methods such as fructosamine (discussed later) should be considered for these patients. Vitamins C and E are reported to falsely lower test results, possibly by inhibiting glycation of hemoglobin. The American Diabetes Association (ADA)

has now endorsed using the HbA$_{1c}$ as a diagnostic test for diabetes (see Table 17–11). A cutoff value of 6.5% was chosen because the risk for retinopathy increases substantially above this value. The advantages of using the HbA$_{1c}$ to diagnose diabetes is that there is no need to fast; it has lower intra-individual variability than the fasting glucose test and the oral glucose tolerance test; and it gives a better picture of glucose control for 2 to 3 months. People with HbA$_{1c}$ levels of 5.7 % to 6.4% should be considered at high risk for developing diabetes (prediabetes). The diagnosis should be confirmed with a repeat HbA$_{1c}$ test, unless the patient is symptomatic with plasma glucose levels more than 200 mg/dL. This test would not be appropriate to use in populations with high prevalence of hemoglobinopathies or in conditions with increased red cell turnover. Also, the testing should be performed using a National Glycohemoglobin Standardization Program certified method and standardized to the Diabetes Control and Complications Trial assay. In the Europeon Union, the tests are standardized to the International Federation of Clinical Chemistry (IFCC) which defines HbA$_{1c}$ as mmol glycated hexapeptide per mol (glycated and nonglycated hexapeptides) and reported as mmol HbA$_{1c}$/mol Hb. The conversion factor for NGSP and IFCC results is: NGSP (USA) result = (0.09148 × IFCC result) + 2.152. Officially there is worldwide consensus that HbA$_{1c}$ should be reported in both NGSP (%) and IFCC (mmol/mol) units.

Serum Fructosamine

Serum fructosamine is formed by nonenzymatic glycosylation of serum proteins (predominantly albumin). Because serum albumin has a much shorter half-life (14-21 days) than hemoglobin, serum fructosamine generally reflects the state of glycemic control for the preceding 2 or 3 weeks. Reductions in serum albumin (eg, nephrotic state or hepatic disease) lower the serum fructosamine value. When abnormal hemoglobins or hemolytic states affect the interpretation of GHb or when a narrower time frame is required, such as for ascertaining glycemic control at the time of conception in a woman with diabetes who has recently become pregnant, serum fructosamine assays offer some advantage. Normal values vary in relation to the serum albumin concentration and are 200 to 285 umol/L when the serum albumin level is 5 g/dL. HbA$_{1c}$ values and fructosamine are highly correlated.

Thus fructosamine levels of 300, 367, and 430 approximate to HbA$_{1c}$ values of 7%, 8%, and 9%, respectively but there is substantial individual variability and caution should be exercised in estimating the likely HbA$_{1c}$ value from the fructosamine measurement.

In most circumstances, glycohemoglobin assays remain the preferred method for assessing long-term glycemic control in patients with diabetes.

Oral Glucose Tolerance Test

It is easy to screen for diabetes using an HbA$_{1c}$ or a fasting plasma glucose level (see Table 17–11). The oral glucose tolerance test, therefore, is mostly performed for research studies or when there is a suspicion of the diagnosis but the fasting plasma glucose is less than 126 mg/dL or the HBA$_{1c}$ level is below 6.5%. The test might

be considered, for example, in a woman with a history of delivering an infant above 9 lb (4.1 kg).

In order to optimize insulin secretion and effectiveness, especially when patients have been on a low-carbohydrate diet, a minimum of 150 to 200 g of carbohydrate per day should be included in the diet for 3 days preceding the test. The patient should eat nothing after midnight prior to the test day. Adults are given 75 g of glucose in 300 mL of water; children are given 1.75 g of glucose per kilogram of ideal body weight. The glucose load is consumed within 5 minutes. The test should be performed in the morning because there is some diurnal variation in oral glucose tolerance and patients should not smoke; drink coffee, tea, or alcohol; or be active during the test.

Blood samples for plasma glucose are obtained at 0 and 120 minutes after ingestion of glucose. A fasting plasma glucose value of 126 mg/dL (7 mmol/L) or higher or a 2-hour value of greater than 200 mg/dL (11.1 mmol/L) is diagnostic of diabetes mellitus (see Table 17–11). An oral glucose tolerance test is normal if the fasting venous plasma glucose value is less than 100 mg/dL (5.6 mmol/L) and the 2-hour value falls below 140 mg/dL (7.8 mmol/L). Patients with 2-hour values of 140 to 199 mg/dL have impaired glucose tolerance. False-positive results may occur in patients who are malnourished at test time, bedridden, or afflicted with an infection or severe emotional stress. Diuretics, oral contraceptives, glucocorticoids, excess thyroxine, phenytoin, nicotinic acid, and some of the psychotropic drugs may also cause false-positive results.

Insulin Levels

Normal immunoreactive insulin levels range from 5 to 20 μU/mL in the fasting state. During an oral glucose tolerance test, they reach 50 to 130 μU/mL at 1 hour, and usually return to levels below 30 μU/mL by 2 hours. Insulin measurements are rarely of clinical usefulness. They are principally used in research studies to determine insulin sensitivity.

The homeostasis model of insulin resistance (HOMA$_{IR}$) estimates insulin sensitivity using the following formula:

$$\text{HOMA}_{IR} \text{ (mmol/L} \times \mu\text{U/mL)} = \text{(fasting glucose [mmol/L]} \times \text{fasting insulin [μU/mL])}/22.5$$

The higher the HOMA$_{IR}$ value the more resistant the individual. Data from the oral glucose tolerance test can also be used to estimate insulin sensitivity. The Matsuda & DeFronzo Insulin Sensitivity Index is calculated as:

$$\text{Insulin sensitivity index (ISI)}$$
$$= 10,000/\text{square root of [fasting glucose} \times \text{fasting insulin]} \times \text{[mean glucose} \times \text{mean insulin during OGTT]}$$

The lower the ISI the more insulin resistant the subject.

Intravenous Glucose Tolerance Test

The intravenous glucose tolerance test (IVGTT) is performed by giving a bolus of 50 g of glucose per 1.7 m^2 body surface area

(or 0.5 g/kg of ideal body weight) as a 25% or 50% solution over 2 to 3 minutes after an overnight fast. Timing begins with injection and samples for plasma glucose determination are obtained from an indwelling needle in the opposite arm at 0, 10, 15, 20, and 30 minutes. The plasma glucose values are plotted on semilogarithmic paper against time. K, a rate constant that reflects the rate of fall of blood glucose in percent per minute, is calculated by determining the time necessary for the glucose concentration to fall by one-half ($t_{1/2}$) and using the following equation:

$$K(\text{glucose}) = \frac{0.693}{t_{1/2}} \times 100$$

The average K value for a nondiabetic patient is approximately 1.72% per minute; this value declines with age but remains above 1.3% per minute. Patients with diabetes almost always have a K value of less than 1% per minute. The disappearance rate reflects the patient's ability to dispose of a glucose load. Perhaps its most widespread present use is to screen siblings at risk for type 1 diabetes to determine if autoimmune destruction of β cells has reduced first phase insulin responses (at 1-5 minutes after the glucose bolus) to levels below the normal lower limit of 40 μU/mL.

The IVGTT has been modified by giving a glucose dose of 0.3 g/kg, with more frequent plasma sampling and extending the test to 3 to 4 hours. Also at 20 minutes, a 5-minute infusion of insulin is given (0.03 U/kg for the subject who is likely to be insulin sensitive and 0.06 U/kg for the likely resistant subject). Plasma glucose is sampled at 3, 4, 5, 6, 8, 10, 14, 19, 22, 25, 27, 30, 40, 50, 60, 80, 100, 140, and 180 minutes. Analysis of the time course of glucose and insulin during this frequently sampled IVGTT (FSIVGTT) allows for measurements of insulin sensitivity (S_i), that is, fractional glucose clearance per unit insulin concentration; the first phase insulin response ($AIR_{glucose}$); and glucose effectiveness (S_G), the ability of glucose itself to enhance its own disappearance independent of any change in insulin. The data analysis requires use of specific software (Minmod).

Lipoproteins in Diabetes

Levels of circulating lipoproteins are dependent on normal levels and action of insulin, just as is the plasma glucose. In type 1 diabetes, moderately deficient control of hyperglycemia is associated with only a slight elevation of low-density lipoprotein (LDL) cholesterol and serum triglycerides and little if any changes in HDL cholesterol. Once the hyperglycemia is corrected, lipoprotein levels are generally normal. However, patients with type 2 diabetes frequently have a dyslipidemia that is characteristic of the insulin resistance syndrome. Its features are a high serum triglyceride level (300-400 mg/dL), a low HDL cholesterol (<30 mg/dL), and a qualitative change in LDL particles producing a smaller dense LDL whose membrane carries supranormal amounts of free cholesterol. Because *low* HDL cholesterol is a major feature predisposing to macrovascular disease, the term *dyslipidemia* has preempted the previous label of *hyperlipidemia*, which mainly described the elevated triglycerides. Measures designed to correct obesity and hyperglycemia, such as exercise, diet, and

hypoglycemic therapy substantially correct the dyslipidemia but most patients require pharmacotherapy. Chapter 19 discusses these matters in detail.

Clinical Trials in Diabetes

Findings of the Diabetes Control and Complications Trial (DCCT) and of the United Kingdom Prospective Diabetes Study (UKPDS) have confirmed the beneficial effects of intensive therapy to achieve improved glycemic control in both type 1 and type 2 diabetes, respectively.

A. The Diabetes Control and Complications Trial (DCCT),

a long-term therapeutic study involving 1441 patients with type 1 diabetes mellitus, reported that "near" normalization of blood glucose resulted in a delay in the onset and a major slowing of the progression of established microvascular and neuropathic complications of diabetes during a follow-up period of up to 10 years. Multiple daily insulin injections (66%) or insulin pumps (34%) were used in the intensively treated group whereas the conventionally treated group used no more than two injections daily. The intensive group achieved a mean HbA_{1c} of 7.2% (normal: <6%) and a mean blood glucose of 155 mg/dL (8.6 mmol/L) while the conventionally treated group averaged an HbA_{1c} of 8.9% and mean blood glucose of 225 mg/dL (12.5 mmol/L). Over mean follow-up of 7 years, there was an approximately 60% reduction in risk between the two groups in regard to diabetic retinopathy, nephropathy, and neuropathy. The intensively treated group also had a nonsignificant reduction in the risk of macrovascular disease of 41% (95% CI, −10 to 68). Intensively treated patients had a threefold greater risk of serious hypoglycemia as well as a greater tendency toward weight gain. However, there were no deaths definitely attributable to hypoglycemia in any persons in the DCCT study, and no evidence of posthypoglycemic cognitive damage was detected.

Subjects participating in the DCCT study were subsequently enrolled in a follow-up observational study (Epidemiology of Diabetes Interventions and Complications [EDIC]). Even though the between group differences in HbA_{1c} narrowed within 4 years, the group assigned to intensive therapy had a lower risk for retinopathy at 4 years and microalbuminuria at 7 to 8 years of post study follow-up. Moreover, by the end of the 11 year follow-up period, the intensive therapy group had significantly reduced risk of any cardiovascular disease events by 42% (95% CI, 9%-23%; $p = 0.02$). Thus it seems that the benefits of good glucose control persist even if control deteriorates at a later date.

The general consensus of the ADA is that intensive insulin therapy associated with comprehensive self-management training should become standard therapy in most patients with type 1 diabetes after the age of puberty. Exceptions include those with advanced renal disease and the elderly, because the detrimental risks of hypoglycemia outweigh the benefit of tight glycemic control in these groups. In children under age 7 years, the risk of developing brain damage from hypoglycemia contraindicates attempts at tight glycemic control, particularly because diabetic complications do not seem to occur until some years after the onset of puberty.

B. The United Kingdom Prospective Diabetes Study (UKPDS) was a multicenter study designed to determine whether the risk of macrovascular or microvascular complications in patients with type 2 diabetes could be reduced by intensive blood glucose control with oral hypoglycemic agents or insulin and whether any particular therapy was better than the others. Patients aged 25 to 65 years who were newly diagnosed with type 2 diabetes were recruited between 1977 and 1991, and a total of 3867 were studied over 10 years. The median age at baseline was 54 years; 44% were overweight (>20% over ideal weight), and baseline HbA_{1c} was 9.1%. Therapies were randomized to include a control group on diet alone and separate groups intensively treated with insulin, chlorpropamide, glyburide, or glipizide. Metformin was included as a randomization option in a subgroup of 342 overweight patients, and—much later in the study—an additional subgroup of both normal-weight and overweight patients, who were responding unsatisfactorily to sulfonylurea therapy, were randomized to either continue on their sulfonylurea therapy alone or to have metformin combined with it.

After the study was initiated, a further modification was made to evaluate whether tight control of blood pressure with stepwise antihypertensive therapy would prevent macrovascular and microvascular complications in 758 hypertensive patients among this UKPDS population—compared with 390 patients whose blood pressure was treated less intensively. The tight control group was randomly assigned to treatment with either an angiotensin-converting enzyme (ACE) inhibitor (captopril) or a beta blocker (atenolol). Both drugs were stepped up to maximum doses of 100 mg/d, and then, if blood pressure remained higher than the target level of less than 150/85 mm Hg, more drugs were added in the following stepwise sequence—a diuretic, slow-release nifedipine, methyldopa, and prazosin—until the target level of tight control was achieved. In the control group, hypertension was conventionally treated to achieve target levels less than 180/105 mm Hg, but these patients were not given either ACE inhibitors or beta blockers.

Intensive glycemic therapy with either sulfonylureas, metformin, combinations of these, or insulin achieved mean HbA_{1c} levels of 7.0%. This level of glycemic control decreased the risk of microvascular complications in comparison with conventional therapy (mostly diet alone), which achieved mean levels of HbA_{1c} of 7.9%. Weight gain occurred in intensively treated patients except when metformin was used as monotherapy. There was a trend toward reduction in cardiovascular events (fatal or nonfatal MI; sudden death) with intensive treatment but this did not reach statistical significance (16% reduction, $p = 0.052$). Hypoglycemic reactions occurred in the intensive treatment groups, but only one death from hypoglycemia was documented over 27,000 patient years of intensive therapy.

Tight control of blood pressure (median value 144/82 mm Hg vs 154/87 mm Hg) substantially reduced the risk of microvascular disease and stroke but not myocardial infarction. The blood pressure lowering had substantially greater impact on microvascular outcomes than that achieved by lowering HbA_{1c} from 7.9% to 7%. More than half of the patients needed two or more drugs for adequate therapy of their hypertension, and there was no demonstrable advantage of ACE inhibitor therapy over beta blockers as regards diabetes endpoints. Use of a calcium channel blocker added to both treatment groups appeared to be safe over the long term in this population with diabetes, despite some controversy in the literature about its safety in such individuals.

The UKPDS researchers, like the DCCT group, performed posttrial monitoring to determine if there were long-term benefits of having been in the intensively treated glucose and blood pressure arms of the study. The between group differences in HbA_{1c} were lost within the first year of follow-up but the reduced risk of development or progression of microvascular complications in the intensively treated group persisted for 10 years (24%, $p = 0.001$). The intensively treated group also had significantly reduced risk for myocardial infarction (15%, $p = 0.01$) and death from any cause (13%, $p = 0.007$) during the follow-up period. The subgroup of overweight or obese subjects who were initially randomized to metformin therapy showed sustained reduction in risk of myocardial infarction and death from any cause in the follow-up period. The between group blood pressure differences disappeared within 2 years of the end of the trial. Unlike the sustained benefits seen with glucose control, there was no sustained benefit from having been in the more tightly controlled blood pressure group. Both blood pressure groups had similar risks for microvascular events and diabetes related endpoints in the follow-up period.

Thus, the follow-up of the UKPDS type 2 diabetes cohort showed that, as in type 1 diabetes, the benefits of good glucose control persist even if control deteriorates at a later date. Blood pressure benefits, however, last only as long as the blood pressure is well controlled.

C. Diabetes Prevention Program (DPP) was a randomized clinical trial in 3234 overweight men and women, aged 25 to 85 years, who showed impaired glucose tolerance. Results from this study indicated that intervention with a low-fat diet and 150 minutes of moderate exercise (equivalent to a brisk walk) per week reduces the risk of progression to type 2 diabetes by 58% as compared with a matched control group. Another arm of this trial demonstrated that use of 850 mg of metformin twice daily reduced the risk of developing type 2 diabetes by 31% but was relatively ineffective in those who were either less obese or in the older age group.

D. The Steno-2 Study This was designed in 1990 to validate the efficacy of targeting multiple concomitant risk factors for both microvascular and macrovascular disorders in type 2 diabetes. A prospective, randomized, open, blinded endpoint design was used in which 160 patients with type 2 diabetes and microalbuminuria were assigned to conventional therapy with their general practitioner or to intensive care at the Steno Diabetes Center. In the intensively treated group, stepwise introduction of lifestyle and pharmacologic interventions was aimed at keeping glycated hemoglobin less than 6.5%, blood pressure less than 130/80 mm Hg, total cholesterol less than 175 mg/dL, and triglycerides less than 150 mg/dL. All subjects in the intensively treated group received ACE inhibitors and if intolerant, an angiotensin II receptor blocker. The lifestyle component of intensive intervention

included reduction in dietary fat intake to less than 30% of total calories, a smoking cessation program, light to moderate exercise, and a daily vitamin-mineral supplement (vitamins C and E and chromium picolinate). Initially, aspirin was given as secondary prevention to patients with a history of ischemic cardiovascular disease, but later all patients received aspirin.

After a mean follow-up of 7.8 years, 44% of patients in the conventional arm and 24% in the intensive multifactorial arm developed cardiovascular events (myocardial infarction, angioplasties, coronary bypass grafts, strokes, amputations, vascular surgical interventions)—a 53% reduction. Rates of nephropathy, retinopathy, and autonomic neuropathy were also lower in the multifactorial intervention arm—61%, 58%, and 63% of rates in the conventional arm, respectively.

The subjects who participated in this trial were subsequently enrolled in an observational follow-up study for an average of 5.5 years. Even though the significant differences in glycemic control and levels of risk factors for cardiovascular disease between the groups had disappeared by the end of the follow-up period, the interventional group continued to have a lower risk for retinal photocoagulation, renal failure, cardiovascular endpoints, and cardiovascular mortality.

The data from the UKPDS and this study thus provide support for guidelines recommending vigorous treatment of concomitant microvascular and cardiovascular risk factors in patients with type 2 diabetes.

E. Accord, Advance, and VADT studies The ACCORD study was a randomized controlled study designed to determine whether normal HbA_{1c} levels would reduce the risk of cardiovascular events in middle-aged or older individuals with type 2 diabetes. About 35% of the 10,251 recruited subjects had established cardiovascular disease at study entry. The intensive arm of the study was discontinued after 3.5 years of follow-up because of more unexplained deaths in the intensive arm when compared to the conventional treatment arm (22%, $p = 0.020$). Analysis of the data at time of discontinuation showed that the intensively treated group (HbA_{1c} 6.4%) had a 10% reduction in cardiovascular event rate compared to the standard treated group (HbA_{1c} 7.5%), but this difference was not statistically significant. The ADVANCE trial randomly assigned 11,140 patients in their 60s with type 2 diabetes to standard or intensive glucose control. The primary outcomes were major macrovascular (nonfatal myocardial infarction or stroke or death from cardiovascular causes) or microvascular events. Thirty-two percent of the subjects had established cardiovascular disease at study entry. After a median follow-up of 5 years, there was a nonsignificant reduction (6%) in major macrovascular event rate in the intensively treated group (HbA_{1c} 6.5%) compared to the standard therapy group (HbA_{1c} 7.3%).

The Veteran Administration Diabetes Trial (VADT) randomly assigned 1791 patients in their 50s and 60s with type 2 diabetes to standard or intensive glucose control. Ninety-seven percent of the subjects were men. The primary outcome was a composite of myocardial infarction, death from cardiovascular causes, congestive heart failure, vascular surgery, inoperable coronary artery disease, and amputation for gangrene. All patients had optimized blood pressure and lipid levels. After a median follow-up of 5.6 years, there was no significant difference in the primary outcome in the intensively treated group (HbA_{1c} 6.9%) compared to the standard therapy (HbA_{1c} 8.4%). Within this larger study, there was an embedded study evaluating the impact of intensive therapy in patients who were categorized as having low, moderate, and high coronary calcium scores on CT scans. Patients with low coronary calcium score showed reduced number of cardiovascular events with intensive therapy.

Thus, the ACCORD, ADVANCE, and VADT results do not provide support for the hypothesis that near-normal glucose control in type 2 diabetes will reduce cardiovascular events. It is, however, important not to over-interpret the results of these three studies. The results do not exclude the possibility that cardiovascular benefits might accrue with longer duration of near-normal glucose control. In the UKPDS, risk reductions for myocardial infarction and death from any cause were only observed during 10 years of post-trial follow-up. Specific subgroups of type 2 diabetes patients may also have different outcomes. The ACCORD, ADVANCE, and VADT studies recruited patients who had diabetes for 8 to 10 years, and a third of them had established cardiovascular disease. Patients in the UKPDS, in contrast, had newly diagnosed diabetes, and only 7.5% had a history of macrovascular disease. It is possible that the benefits of tight glycemic control on macrovascular events are attenuated in patients with longer duration of diabetes or with established vascular disease. Specific therapies used to lower glucose may also affect cardiovascular event rate or mortality. Severe hypoglycemia occurred more frequently in the intensively treated groups of the ACCORD, ADVANCE, and VADT studies; and the ACCORD investigators were not able to exclude undiagnosed hypoglycemia as a potential cause for the increased death rate in their intensive arm group.

A formal meta-analysis performed on the raw trial data from the ACCORD, ADVANCE, VADT, and UKPDS studies found that allocation to more intensive glucose control reduced the risk of MI by 15% (hazard ratio 0.85, 95% CI 0.76-0.94). The benefit appeared to be in patients who did not have preexisting macrovascular disease.

TREATMENT OF DIABETES MELLITUS

Diet

A well-balanced, nutritious diet remains a fundamental element of therapy for diabetes. It is recommended that the macronutrient proportions (carbohydrate, protein, and fat) be individualized based on the patient's eating patterns, preferences and goals. Generally most patients with diabetes consume about 45% of their calories as carbohydrates; 25% to 35% fats; and 10% to 35% proteins. Limiting the carbohydrate intake and substituting some of the calories with monounsaturated fats, such as olive oil, rapeseed (canola) oil, or the oils in nuts and avocados, can lower triglycerides and increase HDL cholesterol. A Mediterranean-style eating pattern (a diet supplemented with walnuts, almonds, hazelnuts, and olive oil) has been shown to improve glycemic control and lower combined endpoints for cardiovascular events and

stroke. Caloric restriction and weight loss is an important goal for the obese patient with type 2 diabetes.

Patients with type 1 diabetes or type 2 diabetes on insulin should be taught *carbohydrate counting*, so they can administer their insulin bolus for each meal based on its carbohydrate content.

The current recommendations for saturated fats and dietary cholesterol intake for people with diabetes are the same as for the general population. Saturated fats should be limited to less than 10% of daily calories and dietary cholesterol intake should be less than 300 mg/d. For those patients with kidney disease, dietary protein should be maintained at the recommended daily allowance of 0.8 g/kg/d. Exchange lists for meal planning can be obtained from the American Diabetes Association and its affiliate associations or from the American Dietetic Association, 216 W. Jackson Blvd., Chicago, IL 60606 (312-899-0040 or http://www.eatright.org).

Special Considerations in Dietary Control

A. Dietary fiber   Plant components such as cellulose, gum, and pectin are indigestible by humans and are termed **dietary fiber**. Insoluble fibers such as cellulose or hemicellulose increase stool bulk and decrease transit time. Soluble fibers such as gums and pectins, found in beans, oatmeal, or apple skin, can delay glucose absorption and so diminish postprandial hyperglycemia. Although the ADA diet does not require insoluble fiber supplements such as added bran, it recommends foods such as oatmeal, cereals, and beans with relatively high soluble fiber content as stable components of the diet in patients with diabetes. High soluble fiber content in the diet may also have a favorable effect on blood cholesterol levels.

B. Glycemic index   Quantitation of the relative glycemic contribution of different carbohydrate foods has formed the basis of a **glycemic index** in which the area of blood glucose (plotted on a graph) generated over a 3-hour period following ingestion of a test food containing 50 g of carbohydrate is compared with the area plotted after giving a similar quantity of reference food such as glucose or white bread:

$$\frac{\text{Blood glucose area of test food}}{\text{Blood glucose area of reference food}} \times 100$$

White bread is preferred to glucose as a reference standard because it is more palatable and has less tendency to slow gastric emptying by high tonicity, as happens when glucose solution is used. Eating foods with low glycemic index will result in lower glucose levels after meals. Low glycemic index foods have values of 55 or less and include many fruits (apples, oranges) and vegetables, grainy breads, pasta, legumes, milk, and yoghurt. High glycemic index foods have values of 70 and over and include baked potato, white bread, and most white rice. Glycemic index is lowered by the presence of fats and protein when the food is consumed in a mixed meal. Cooking methods can also affect the glycemic index—thus mashed potatoes have a higher glycemic index than baked potato.

Since you have to have 50 g of available carbohydrate to measure the glycemic index, you cannot assign glycemic indices to foods which have very little carbohydrate. Even though it may not be possible to accurately predict the impact of the glycemic index of a particular food in the context of a meal, it is still reasonable to choose foods with low glycemic index.

C. Sweeteners   The nonnutritive sweetener **saccharin** is widely used as a sugar substitute (sweet and low). **Aspartame** (NutraSweet) consists of two major amino acids, aspartic acid, and phenylalanine, which combine to produce a nutritive sweetener 180 times as sweet as sucrose. A major limitation is its heat lability, which precludes its use in baking or cooking. Sucralose (Splenda) and acesulfame potassium (Sunett, Sweet One, DiabetiSweet) are two other nonnutritive sweeteners that are heat stable and can be used in cooking and baking.

Fructose represents a *natural* sugar substance that is a highly effective sweetener. It induces only slight increases in plasma glucose levels and does not require insulin for its utilization. However, because of potential adverse effects of large amounts of fructose (up to 20% of total calories) on raising serum cholesterol, triglycerides, and LDL cholesterol, it does not have any advantage as a sweetening agent in the diabetic diet. This does not preclude, however, ingestion of fructose-containing fruits and vegetables or fructose-sweetened foods in moderation.

Sugar alcohols, also known as polyols or polyalcohol, are commonly used as sweeteners and bulking agents. They occur naturally in a variety of fruits and vegetables but are also commercially made from sucrose, glucose, and starch. Examples are sorbitol, xylitol, mannitol, lactitol, isomalt, maltitol, and hydrogenated starch hydrolysates (HSH). They are not as easily absorbed as sugar, so they do not raise blood glucose levels as much as conventional sugars. Therefore, sugar alcohols are often used in food products that are labeled as *sugar free*, such as chewing gum, lozenges, hard candy, and sugar-free ice cream. However, if consumed in large quantities, they will raise blood glucose and can also cause bloating and diarrhea.

D. Fish oils and other oils   Omega-3 fatty acids in high doses have been shown to lower plasma triglycerides and VLDL cholesterol. They may also reduce platelet aggregation. In the Lyon Diet Heart Study in nondiabetic patients, a high intake of α-linolenic acid was beneficial in secondary prevention of coronary heart disease. This diet, which is rich in vegetables and fruits, also supplies a high intake of natural antioxidants. There is limited clinical information on the use of these oils in patients with diabetes.

AGENTS FOR THE TREATMENT OF HYPERGLYCEMIA

The drugs for treating type 2 diabetes (Table 17–12), other than insulin, fall into several categories. (1) **Drugs that act on the sulfonylurea receptor complex of the β cell**. Sulfonylureas remain the most widely prescribed drugs for treating hyperglycemia.

TABLE 17–12 Drugs for treatment of type 2 diabetes.

Drug	Tablet Size	Daily Dose	Duration of Action (h)
Sulfonylureas			
Acetohexamide (Dymelor)	250 and 500 mg	0.25-1.5 g as single dose or in two divided doses	8-24
Chlorpropamide (Diabinese)	100 and 250 mg	0.1-0.5 g as a single dose	24-72
Gliclazide (not available in United States)	80 mg	40-80 mg as single dose; 160-320 mg as divided dose	12
Glimepiride (Amaryl)	1, 2, and 4 mg	1-4 mg as single dose. 8 mg once a day is maximal dose	Up to 24
Glipizide			
(Glucotrol)	5 and 10 mg	2.5-40 mg as single dose or in two divided doses on an empty stomach	6-12
(Glucotrol XL)	2.5, 5, and 10 mg	Up to 20 mg daily as a single dose	Up to 24
Glyburide			
(DiaBeta, Micronase)	1.25, 2.5, and 5 mg	1.25-20 mg as single dose or in two divided doses	Up to 24
Glynase	1.5, 3, and 6 mg	1.5-19 mg as a single dose or in two divided doses	Up to 24
Tolazamide (Tolinase)	100, 250, and 500 mg	0.1-1 g as single dose or in two divided doses	Up to 24
Tolbutamide (Orinase)	250 and 500 mg	0.5-2 g in two or three divided doses	6-12
Meglitinide Analogs			
Repaglinide (Prandin)	0.5, 1, and 2 mg	0.5 to 4 mg three times a day before meals	3
Mitiglinide (available in Japan)	5 and 10 mg	5 or 10 mg three times a day before meals	2
D-Phenylalanine			
Nateglinide (Starlix)	60 and 120 mg	60 or 120 mg three times a day before meals	1.5
Biguanides			
Metformin (Glucophage)	500, 850, and 1000 mg	1-2.5 g; one tablet with meals two or thee times daily	7-12
Extended-release metformin (Glucophage XR)	500 and 750 mg	500-2000 mg once a day	Up to 24
Thiazolidinediones			
Pioglitazone (Actos)	15, 30, and 45 mg	15-45 mg daily	Up to 24
Rosiglitazone (Avandia)	2, 4, and 8 mg	4-8 mg daily (can be divided)	Up to 24
α-Glucosidase Inhibitors			
Acarbose (Precose)	50 and 100 mg	75-300 mg in three divided doses with first bite of food	4
Miglitol (Glyset)	25, 50, and 100 mg	75-300 mg in three divided doses with first bite of food	4
GLP-1 Receptor Agonists			
Albiglutide (Tanzeum)	30, 50 mg single dose pen (powder)	Mix with diluent and inject subcutaneously. 30 mg is usual dose. Can increase dose to 50 mg if necessary	1 week
Dulaglutide (Trulicity)	0.75, 1.5 mg single dose pen or prefilled syringe	0.75 mg subcutaneously. Dose can be increased to 1.5 mg if necessary	1 week
Exenatide (Byetta)	5 and 10 μg	5 μg by subcutaneous injection within 1 hour of breakfast and dinner. Increase to 10 μg twice a day after about a month. Do not use if calculated creatinine clearance is <30 mL/min	6
Exenatide long-acting release (Bydureon, Byetta LAR)	2 mg powder	Suspend in provided diluent and inject subcutaneously	1 week
Liraglutide (Victoza)	0.6, 1.2, and 1.8 mg	0.6 mg by subcutaneous injection once a day starting dose. Increase to 1.2 mg after a week if no adverse reactions. Dose can be further increased to 1.8 mg if necessary	24
Lixisenatide (Adlyxin)	10 and 10 μg	10 μg by subcutaneous injection daily. Increase to 20 μg after two weeks	24
DPP-4 Inhibitors			
Alogliptin (Nesina)	6.25, 12.5, and 25 mg	25 mg once daily; dose is 12.5 mg daily if calculated creatinine clearance is 30–59 mL/min and 6.25 mg daily if clearance <30 mL/min	24
Linagliptin (Tradjenta)	5 mg	5 mg daily	24
Saxagliptin	2.5 and 5 mg	2.5 or 5 mg once a day. Use 2.5 mg dose if calculated creatinine clearance is ≤50 mL/min or if also taking strong CYP 3A4/5 inhibitors like ketoconazole	24
Sitagliptin	25, 50, and 100 mg	100 mg orally once a day. Reduce dose to 50 mg if calculated creatinine clearance is 30-50 mL/min. Give 25 mg daily if creatinine clearance <30 mL/min	24

(Continued)

TABLE 17–12 Drugs for treatment of type 2 diabetes. (*Continued*)

Drug	Tablet Size	Daily Dose	Duration of Action (h)
Vildagliptin (not available in United States)	50 mg	50 mg once or twice a day. Contraindicated in patients with calculated creatinine clearance ≤60 mL/min or AST/ALT three times upper limit of normal	24
SGLT2 Inhibitors Canagliflozin (Invokana)	100 and 300 mg	100 mg daily is usual dose. Do not use if eGFR <45 mL/min/1.72 m^2 300 mg dose can be used if normal eGFR	24
Dapagliflozin (Farxiga)	5 and 10 mg	10 mg daily. Use 5 mg dose in hepatic failure	24
Empagliflozin (Jardiance)	10 and 25 mg	10 mg daily. 25 mg can be used if necessary	24
Others Bromocriptine (Cycloset)	0.8 mg	0.8 mg daily. Increase weekly by 1 tablet until maximal tolerated dose of 1.6-4.8 mg daily	24
Colesevelam (Welchol)	625 mg	3 tablets twice a day	24
Pramlintide	5-mL vial containing 0.6 mg/mL	For insulin-treated type 2 patients, start at 60-μg dose three times a day (10 U on U100 insulin syringe) and increase to 120 μg three times a day (20 U) if patient has no nausea for 3-7 days. Give immediately before meal. For type 1 patients, start at 15 μg three times a day (2.5 U on U100 insulin syringe) and increase by increments of 15 μg to a maximum of 60 μg three times a day as tolerated. Lower insulin dose by 50% on initiation of therapy to avoid hypoglycemia	2

The meglitinide analog repaglinide and the D-phenylalanine derivative nateglinide also bind the sulfonylurea receptor and stimulate insulin secretion. (2) **Drugs that principally lower glucose levels by their actions on liver, skeletal muscle, or adipose tissue**. Metformin works primarily in the liver. The peroxisome proliferator-activated receptor agonists (PPARs), rosiglitazone, and pioglitazone appear to have their main effects on skeletal muscle and adipose tissue. (3) **Drugs that principally affect absorption of glucose**. The α-glucosidase inhibitors acarbose and miglitol are currently available drugs in this class. (4) **Drugs that mimic incretin effects or prolong incretin action**. The GLP-1 receptor agonists and the DPP-4 inhibitors fall into this category. (5) **Drugs that inhibit reabsorption of filtered glucose in the kidney**. The sodium-glucose cotransporter inhibitors are such agents. (6) **Other drugs**. Pramlintide, lowers glucose by suppressing glucagon and slowing gastric emptying. The mechanisms by which bromocriptine and colesevelam lower glucose levels have not been defined.

1. DRUGS THAT ACT ON THE SULFONYLUREA RECEPTOR COMPLEX

Sulfonylureas

The sulfonylureas contain a sulfonic acid-urea nucleus that can be modified by chemical substitutions to produce agents that have similar qualitative actions but differ widely in potency. They bind the ATP-sensitive potassium channels (K_{ATP}) on the surface of pancreatic β cells, resulting in closure of the channel and depolarization of the β cell. This depolarized state permits calcium to enter the cell and actively promote insulin release (see Figure 17–5).

In β cells, the ATP-sensitive channels contain four copies each of two subunits, the regulatory subunit SUR1, which binds ATP, ADP, and sulfonylureas, and the potassium channel subunit Kir6.2. K_{ATP} channels composed of the same subunits are found in α cells, GLP-1 secreting intestinal L-cells, and the brain. The SUR1/Kir6.2 complexes are opened by diazoxide and closed by sulfonylureas at low concentrations (IC_{50} about 1 nM for glyburide). Inactivating mutations in SUR1 or Kir6.2 cause persistent depolarization of the β cells and have been identified in patients with hyperinsulinemic hypoglycemia of infancy (see Chapter 18). Activating mutations in SUR1 or Kir6.2 prevent depolarization of the β cells and have been identified in patients with neonatal diabetes (see section on Neonatal Diabetes earlier).

K_{ATP} channels with different subunit combinations are found in other tissues. The combination of SUR2A/Kir6.2 is found in cardiac and skeletal muscle, and SUR2B/Kir6.1 in vascular smooth muscle. Channel configurations containing SUR2 subunits are insensitive to diazoxide but sensitive to other potassium channel agonists such as pinacidil and cromakalim. Certain channel *closers* have much higher affinity for SUR1-containing channels than SUR2-containing channels (the sulfonylureas tolbutamide and gliclazide and the meglitinides nateglinide and mitiglinide), while others have similar affinities for both of them (glyburide, glimepiride, and repaglinide). It remains uncertain whether the different affinities of these drugs for the two classes of receptors have clinical relevance.

Sulfonylureas are not indicated for use in type 1 diabetes patients since these drugs require functioning pancreatic β cells to produce their effect on blood glucose. These drugs are used in patients with type 2 diabetes, in whom acute administration improves the early phase of insulin release that is refractory to acute glucose stimulation. Sulfonylureas are metabolized by the

liver and apart from acetohexamide, whose metabolite is more active than the parent compound, the metabolites of all the other sulfonylureas are weakly active or inactive. The metabolites are excreted by the kidney and, in the case of the second-generation sulfonylureas, partly excreted in the bile. Sulfonylureas are generally contraindicated in patients with severe liver or kidney impairment. Idiosyncratic reactions are rare, with skin rashes or hematologic toxicity (leukopenia, thrombocytopenia) occurring in less than 0.1% of users.

A. First-generation sulfonylureas (tolbutamide, tolazamide, acetohexamide, and chlorpropamide)
Tolbutamide is supplied in tablets of 500 mg. It is rapidly oxidized in the liver to an inactive form. Because its duration of effect is short (6-10 hours), it is usually administered in divided doses (eg, 500 mg before each meal and at bedtime). The usual daily dose is 1.5 to 3 g; some patients, however, require only 250 to 500 mg daily. Acute toxic reactions such as skin rashes are rare. Because of its short duration of action, which is independent of renal function, tolbutamide is probably the safest agent to use in elderly patients, in whom hypoglycemia would be a particularly serious risk. Prolonged hypoglycemia has been reported rarely, mainly in patients receiving certain drugs (eg, warfarin, phenylbutazone, or sulfonamides) that compete with sulfonylureas for hepatic oxidation, resulting in maintenance of high levels of unmetabolized active sulfonylureas in the circulation.

Tolazamide, acetohexamide, and chlorpropamide are rarely used. Chlorpropamide has a prolonged biologic effect, and severe hypoglycemia can occur especially in the elderly, because their renal clearance declines with aging. Its other side effects include alcohol-induced flushing and hyponatremia due to its effect on vasopressin secretion and action.

Second-generation sulfonylureas (glyburide, glipizide, gliclazide, and glimepiride)
These agents have similar chemical structures, with cyclic carbon rings at each end of the sulfonylurea nucleus; this causes them to be highly potent (100- to 200-fold more potent than tolbutamide). The drugs should be used with caution in patients with cardiovascular disease as well as in elderly patients, in whom hypoglycemia would be especially dangerous.

a. Glyburide (glibenclamide)—Glyburide is supplied in tablets containing 1.25, 2.5, and 5 mg. The usual starting dose is 2.5 mg/d, and the average maintenance dose is 5 to 10 mg/d given as a single morning dose. If patients are going to respond to glyburide, they generally do so at doses of 10 mg/d or less, given once daily. If they fail to respond to 10 mg/d, it is uncommon for an increase in dosage to result in improved glycemic control. Maintenance doses higher than 20 mg/d are not recommended and may even worsen hyperglycemia. Glyburide is metabolized in the liver into products with such low hypoglycemic activity that they are considered clinically unimportant unless renal excretion is compromised. Although assays specific for the unmetabolized compound suggest a plasma half-life of only 1 to 2 hours, the biologic effects of glyburide clearly persist for 24 hours after a single morning dose in diabetic patients. Glyburide is unique among sulfonylureas in that it not only binds to the pancreatic β cell membrane sulfonylurea receptor but also

becomes sequestered within the β cell. This may also contribute to its prolonged biologic effect despite its relatively short circulating half-life.

A formulation of micronized glyburide, which apparently increases its bioavailability, is available in easy to divide tablet sizes of 1.5, 3, and 6 mg.

Glyburide has few adverse effects other than its potential for causing hypoglycemia. It should not be used in patients with liver failure and renal failure because of the risk of hypoglycemia. Elderly patients are at particular risk of hypoglycemia even with relatively small daily doses. Glyburide does not cause water retention, as chlorpropamide does, and even slightly enhances free water clearance.

b. Glipizide (glydiazinamide)—Glipizide is supplied in tablets containing 5 and 10 mg. For maximum effect in reducing postprandial hyperglycemia, this agent should be ingested 30 minutes before breakfast, because rapid absorption is delayed when the drug is taken with food. The recommended starting dose is 5 mg daily but 2.5-mg dose may be sufficient in elderly patients with early diabetes. The dose can gradually be increased by 2.5 or 5-mg increments. Although as much as 15 mg can be given as a single daily dose before breakfast, most patients do better with divided dosing, taking 5 mg before each meal or taking 10 mg before breakfast and before dinner. The maximum recommended dose is 40 mg/d, although doses above 20 mg probably provide little additional benefit in poor responders. At least 90% of glipizide is metabolized in the liver to inactive products, and only a small fraction is excreted unchanged in the urine. Glipizide therapy is contraindicated in patients with liver failure. Because of its lower potency and shorter half-life, it is preferable to glyburide in elderly patients and for those patients with renal impairment.

Glipizide is also available as a slow-release preparation (Glucotrol-XL, 2.5, 5, and 10-mg tablets). The medication is enclosed in a nonabsorbable shell that contains an osmotic compartment that expands slowly, thereby slowly pumping out the glipizide in a sustained manner. It provides extended release during transit through the gastrointestinal tract, with greater effectiveness in lowering of prebreakfast hyperglycemia than the shorter duration, immediate-release standard glipizide tablets. However, this formulation appears to have sacrificed glipizide's reduced propensity for severe hypoglycemia compared with longer acting glyburide without showing any demonstrable therapeutic advantages over glyburide.

c. Gliclazide (not available in the United States)—This drug is another intermediate duration sulfonylurea with a duration of action of about 12 hours. It is available as 80-mg tablets. The recommended starting dose is 40 to 80 mg/d with a maximum dose of 320 mg. Doses of 160 mg and above are given as divided doses before breakfast and dinner. The drug is metabolized by the liver, and the metabolites and conjugates have no hypoglycemic effect. An extended-release preparation is also available.

d. Glimepiride—This sulfonylurea is supplied in tablets containing 1, 2, and 4 mg. It has a long duration of effect with a half-life of 5 hours, allowing once-daily administration. Glimepiride achieves blood glucose lowering with the lowest dose of any sulfonylurea compound. A single daily dose of 1 mg/d has been shown to be effective, and the maximal recommended dose is 8 mg. It is completely metabolized by the liver to relatively inactive metabolic products.

Meglitinide Analogs

Repaglinide is supplied as 0.5, 1, and 2-mg tablets. Its structure is similar to that of glyburide but lacks the sulfonic acid-urea moiety. It also acts by binding to the sulfonylurea receptor and closing the ATP-sensitive potassium channel. It is rapidly absorbed from the intestine and then undergoes complete metabolism in the liver to inactive biliary products, giving it a plasma half-life of less than 1 hour. The drug therefore causes a brief but rapid pulse of insulin. The starting dose is 0.5 mg three times a day 15 minutes before each meal. The dose can be titrated to a maximum daily dose of 16 mg. Hypoglycemia is the main side effect. In clinical trials, when the drug was compared with glyburide, a long-acting sulfonylurea, there was a trend toward less hypoglycemia. Like the sulfonylureas, it causes weight gain. Metabolism is by the cytochrome P450 3A4 isoenzyme. Other drugs that induce or inhibit this isoenzyme may increase or inhibit the metabolism of repaglinide. The drug may be useful in patients with renal impairment or in the elderly.

Mitiglinide is a benzylsuccinic acid derivative that is very similar to repaglinide in its clinical effects. It binds to the sulfonylurea receptor causing a brief pulse of insulin. It is given as a 5- or 10-mg dose just before a meal and reduces the postprandial rise in blood glucose. It has been approved for use in Japan.

δ-Phenylalanine Derivative

Nateglinide is supplied in tablets of 60 and 120 mg. This drug binds the sulfonylurea receptor and closes the ATP-sensitive potassium channel. The drug is rapidly absorbed from the intestine, reaching peak plasma levels within 1 hour. It is metabolized in the liver and has a plasma half-life of about 1.5 hours. Like repaglinide, it causes a brief rapid pulse of insulin, and when given before a meal, it reduces the postprandial rise in blood glucose. The 60-mg dose is used in patients with mild elevations in HbA_{1c}. For most patients, the recommended starting and maintenance dosage is 120 mg three times a day before meals. Like the other insulin secretagogues, its main side effects are hypoglycemia and weight gain.

2. DRUGS THAT ACT ON INSULIN TARGET TISSUES

Metformin

Metformin (1,1-dimethylbiguanide hydrochloride) is used, either alone or in conjunction with other oral agents or insulin, in the treatment of patients with type 2 diabetes.

Metformin acts primarily through AMPK (see Figure 17–8), which it activates by uncoupling mitochondrial oxidative phosphorylation and increasing cellular AMP levels. Metformin's therapeutic effects primarily derive from its effects on the liver, where increased AMPK activity reduces hepatic gluconeogenesis and lipogenesis. Metformin is a substrate for organic cation transporter 1, which is abundantly expressed in hepatocytes and in the gut.

Metformin is the first-line therapy for patients with type 2 diabetes. It is ineffective in patients with type 1 diabetes. The current recommendation is to start this drug at diagnosis. A side benefit of metformin therapy is its tendency to improve both fasting and postprandial hyperglycemia and hypertriglyceridemia in obese patients with diabetes without the weight gain associated with insulin or sulfonylurea therapy. It is dispensed as 500-, 850-, and 1000-mg tablets. A 500- and 750-mg extended-release preparation is also available. Eighty-five percent of the maximal glucose-lowering effect is achieved by a daily dose of 1500 mg, and there is little benefit from giving more than 2000 mg daily. It is important to begin with a low dose and increase the dosage very gradually in divided doses—taken with meals—to reduce minor gastrointestinal upsets. A common schedule would be one 500-mg tablet three times a day with meals or one 850-mg or 1000-mg tablet twice daily at breakfast and dinner. The maximum recommended dose is 850 mg three times a day. Up to 2000 mg of the extended-release preparation can be given once a day.

The most frequent side effects of metformin are gastrointestinal symptoms (anorexia, nausea, vomiting, abdominal discomfort, diarrhea), which occur in up to 20% of patients. These effects are dose-related, tend to occur at onset of therapy, and often are transient. However, in 3% to 5% of patients, therapy may have to be discontinued because of persistent diarrheal discomfort. In a retrospective analysis, it has been reported that patients who switched from immediate-release metformin to a comparable dose of extended-release metformin experienced fewer gastrointestinal side effects.

Absorption of vitamin B_{12} appears to be reduced during chronic metformin therapy but usually the vitamin B_{12} levels remain in the normal range. Screening of serum vitamin B_{12} levels should be considered if the patient develops a macrocytic anemia or if the patient develops peripheral neuropathy symptoms. Screening may also be considered in patients who have been on the drug for many years. Increased intake of dietary calcium may prevent the metformin-induced B_{12} malabsorption. If pernicious anemia is suspected then measurements of methylmalonate, homocysteine and antibodies to intrinsic factor should be performed. Hypoglycemia does not occur with therapeutic doses of metformin, which permits its description as a euglycemic or antihyperglycemic drug rather than an oral hypoglycemic agent. Dermatologic or hematologic toxicity is rare.

Metformin both increases lactate production by uncoupling mitochondrial oxidative phosphorylation, especially in the gut, and reduces lactate removal by the liver by blocking gluconeogenesis. At therapeutic doses of metformin, serum lactate levels rise only minimally if at all, since other organs such as the kidney can remove the slight excess. However, if tissue hypoxia occurs, the metformin-treated patient is at higher risk for lactic acidosis due to compromised lactate removal. Metformin has a half-life of 1.5 to 3 hours, does not bind to plasma proteins, and is excreted unchanged by the kidneys. With renal failure, plasma levels of metformin can rise far above the therapeutic range and increase lactate production and block hepatic uptake sufficiently to provoke lactic acidosis even in the absence of other causes of increased lactic acid production. Metformin can be safely used in patients with estimated glomerular filtration rate (eGFR) between 45 and 60 mL/min/1.73 m^2. It can be used cautiously in patients with

eGFR between 30 and 45 mL/min/1.73 m^2. It is contraindicated if the eGFR is less than 30 mL/min/1.73 m^2. Acute kidney failure can occur rarely in certain patients receiving radiocontrast agents. Metformin therapy should therefore be temporarily halted on the day of radiocontrast administration and restarted a day or two later after confirmation that renal function has not deteriorated. Renal function should be checked at least annually in patients on metformin therapy; and lower doses should be used in the elderly who may have limited renal reserve and in those with eGFR between 30 and 45 mL/min/1.73 m^2.

Alcoholics and patients with liver failure should not take metformin—lactic acidosis can occur because of impaired hepatic clearance of lactate. The drug is relatively contraindicated in patients with cardiorespiratory insufficiency, because they have a propensity to develop hypoxia that would aggravate the lactic acid production already occurring from metformin therapy.

Peroxisome Proliferator–Activated Receptor Agonists

Thiazolidinediones are insulin sensitizers exerting their antidiabetic effects through the activation of PPARγ (see discussion on PPAR nuclear receptors in insulin action and insulin resistance earlier). Observed effects of thiazolidinediones include increased GLUT expression (GLUT 1 and GLUT 4); decreased free fatty acid levels; decreased hepatic glucose output; increased adiponectin and decreased resistin release from adipocytes; and increased differentiation of preadipocytes into adipocytes. They have also been demonstrated to decrease levels of plasminogen activator inhibitor type 1, matrix metalloproteinase 9, C-reactive protein, and interleukin-6. Like the biguanides, this class of drugs does not cause hypoglycemia.

Rosiglitazone and **pioglitazone** are available for clinical use. Both are effective as monotherapy and in combination with sulfonylureas, metformin, or insulin. When used as monotherapy, these drugs lower HbA$_{1c}$ by about 1 or 2 percentage points. When used in combination with insulin, they can result in a 30% to 50% reduction in insulin dosage, and some patients can come off insulin completely. The dosage of rosiglitazone is 4 to 8 mg daily and of pioglitazone 15 to 45 mg daily, and the drugs do not have to be taken with food. Rosiglitazone is primarily metabolized by the CYP 2C8 isoenzyme and pioglitazone is metabolized by CYP 2C8 and CYP 3A4.

The combination of a thiazolidinedione and metformin has the advantage of not causing hypoglycemia. Patients inadequately managed on sulfonylureas can do well on a combination of sulfonylurea and rosiglitazone or pioglitazone. About 25% of patients in clinical trials fail to respond to these drugs, presumably because they are significantly insulinopenic.

In addition to glucose-lowering, the thiazolidinediones have effects on lipids and other cardiovascular risk factors. Rosiglitazone therapy is associated with increases in total cholesterol, LDL cholesterol (15%), and HDL cholesterol (10%). There is a reduction in free fatty acids of about 8% to 15%. The changes in triglycerides were generally not different from placebo. Pioglitazone in clinical trials lowered triglycerides (9%) and increased

HDL cholesterol (15%) but did not cause a consistent change in total cholesterol and LDL cholesterol levels. A prospective randomized comparison of the metabolic effects of pioglitazone and rosiglitazone in patients showed similar effects on HbA$_{1c}$ and weight gain. Pioglitazone-treated subjects, however, had lower total cholesterol, LDL cholesterol, and triglyceride levels when compared with rosiglitazone. Small prospective studies have also shown that treatment with these drugs leads to improvement of biochemical and histological features of nonalcoholic fatty liver disease. The thiazolidinediones also may limit vascular smooth muscle proliferation after injury and there are reports that pioglitazone can reduce neointimal proliferation after coronary stent placement. Also, in one double-blind, placebo-controlled study, rosiglitazone was shown to be associated with a decrease in the ratio of urinary albumin to creatinine.

Safety concerns and some troublesome side effects have emerged over this class of drugs that limit their use. A meta-analysis of 42 randomized clinical trials with rosiglitazone suggested that this drug increases the risk of angina pectoris or myocardial infarction. A meta-analysis of clinical trials with pioglitazone did not show a similar finding. Although conclusive data are lacking, the European Medicines Agency suspended the use of rosiglitazone in Europe. In the United States, the FDA established a restricted distribution program for rosiglitazone. A subsequent large prospective clinical trial (the RECORD study) failed to confirm the meta-analysis finding and the restrictions were lifted in the United States.

Edema occurs in about 3% to 4% of patients receiving monotherapy with rosiglitazone or pioglitazone. The edema occurs more frequently (10%-15%) in patients receiving concomitant insulin therapy and may result in congestive heart failure. The drugs are contraindicated in diabetic individuals with New York Heart Association class III or IV heart failure. Thiazolidinediones have also rarely been reported as being associated with new onset or worsening macular edema which resolved or improved once the drug was discontinued.

In experimental animals, rosiglitazone stimulates bone marrow adipogenesis at the expense of osteoblastogenesis resulting in a decrease in bone mineral density. An increase in fracture risk in women (but not men) has been reported with both rosiglitazone and pioglitazone. The fracture risk is in the range of 1.9 per 100 patient years with the thiazolidinedione. In at least one study of rosiglitazone, the fracture risk was increased in premenopausal as well as postmenopausal women. Other side effects include anemia, which occurs in 4% of patients treated with these drugs; it may be due to a dilutional effect of increased plasma volume rather than a reduction in red cell mass. Weight gain occurs, especially when the drug is combined with a sulfonylurea or insulin. Some of the weight gain is fluid retention, but there is also an increase in total fat mass. In preclinical studies with pioglitazone, bladder tumors were observed in male rats receiving clinically relevant doses of the medication. Initial clinical reports indicated that this might also be true in humans. A 10-year observational cohort study of patients taking pioglitazone, however, failed to find an association with bladder cancer. A large multipopulation pooled analysis (1.01 million persons over 5.9 million person years)

also failed to find an association between cumulative exposure to pioglitazone or rosiglitazone and incidence of bladder cancer. Another population based study generating 689,616 person years of follow-up did find that pioglitazone but not rosiglitazone was associated with an increased risk of bladder cancer—hazard ratio 1.63 for pioglitazone, 95% confidence intervals 1.22 to 2.19. Troglitazone, the first drug in this class, was withdrawn from clinical use because of drug-associated fatal liver failure. The two currently available agents, rosiglitazone and pioglitazone, have thus far not caused hepatotoxicity. The FDA has, however, recommended that patients should not initiate drug therapy if there is clinical evidence of active liver disease or pretreatment elevation of the alanine aminotransferase (ALT) level that is 2.5 times greater than the upper limit of normal. Obviously, caution should be used in initiation of therapy in patients with even mild ALT elevations. Liver biochemical tests should be performed prior to initiation of treatment and periodically thereafter.

3. DRUGS THAT AFFECT GLUCOSE ABSORPTION

Alpha-Glucosidase Inhibitors

Drugs of this family are competitive inhibitors of intestinal brush border α glucosidases. Two of these drugs, acarbose and miglitol, are available for clinical use in the United States. Voglibose, another α glucosidase inhibitor, is available in Japan, Korea, and India. Acarbose and miglitol are potent inhibitors of glucoamylase, α-amylase, and sucrase. They are less effective on isomaltase and are ineffective on trehalase or lactase. Acarbose binds 1000 times more avidly to the intestinal disaccharidases than do products of carbohydrate digestion or sucrose. A fundamental difference exists between acarbose and miglitol in their absorption. Acarbose has the molecular mass and structural features of a tetrasaccharide, and very little (~2%) crosses the microvillar membrane. Miglitol, however, is structurally similar to glucose and is absorbable. Both drugs delay the absorption of carbohydrates and reduce postprandial glycemic excursion.

a. **Acarbose** is available as 50- and 100-mg tablets. The recommended starting dose is 50 mg twice daily, gradually increasing to 100 mg three times daily. For maximal benefit on postprandial hyperglycemia, acarbose should be given with the first mouthful of food ingested. In diabetic patients it reduces postprandial hyperglycemia by 30% to 50%, and its overall effect is to lower the HbA$_{1c}$ by 0.5% to 1%. The principal adverse effect, seen in 20% to 30% of patients, is flatulence. This is caused by undigested carbohydrate reaching the lower bowel, where gases are produced by bacterial flora. In 3% of cases, troublesome diarrhea occurs. This gastrointestinal discomfort tends to discourage excessive carbohydrate consumption and promotes improved compliance of patients with type 2 diabetes with their diet prescriptions.

When acarbose is given alone, there is no risk of hypoglycemia. However, if combined with insulin or sulfonylureas, it may increase risk of hypoglycemia from these agents. A slight rise in hepatic aminotransferases has been noted in clinical trials (5% vs 2% in placebo controls, and

particularly with doses >300 mg/d). This generally returns to normal on stopping this drug. In the UKPDS, approximately 2000 patients on diet, sulfonylurea, metformin, or insulin therapy were randomized to acarbose or placebo therapy. By 3 years, 60% of the patients had discontinued the drug, mostly because of gastrointestinal symptoms. In the 40% of patients who remained on the drug, acarbose was associated with a 0.5% lowering of HbA$_{1c}$ compared with placebo.

b. **Miglitol** is similar to acarbose in terms of its clinical effects. It is indicated for use in diet- or sulfonylurea-treated patients with type 2 diabetes. Therapy is initiated at the lowest effective dosage of 25 mg three times a day. The usual maintenance dose is 50 mg three times a day, although some patients may benefit from increasing the dose to 100 mg three times a day. Gastrointestinal side effects occur as with acarbose. The drug is not metabolized and is excreted unchanged by the kidney. Theoretically, absorbable α-glucosidase inhibitors could induce a deficiency of one or more of α-glucosidases involved in cellular glycogen metabolism and biosynthesis of glycoproteins. This does not occur in practice because—unlike the intestinal mucosa, which is exposed to a high concentration of the drug—circulating plasma levels are 200-fold to 1000-fold lower than those needed to inhibit intracellular α glucosidases. Miglitol should not be used in renal failure because its clearance is impaired in this setting.

4. INCRETINS

The gut makes several incretins, gut hormones that amplify postprandial insulin secretion, including glucagon-like peptide-1 (GLP-1, see Figure 17–9) and glucose-dependent insulinotropic polypeptide (GIP). Therapeutic drugs in this class include **GLP-1 receptor agonists** and **dipeptidyl peptidase 4** (DPP-4) inhibitors, which increase levels of both GLP-1 and GIP.

When GLP-1 is infused in patients with type 2 diabetes, it stimulates insulin secretion and lowers glucose levels. GLP-1, unlike the sulfonylureas, has only a modest insulin stimulatory effect under normoglycemic conditions. This means that GLP-1 administration has a lower risk of causing hypoglycemia than the sulfonylureas. GLP-1, in addition to its insulin stimulatory effect, also has a number of other pancreatic and extrapancreatic effects (see Table 17–4).

GLP-1 Receptor Agonists

GLP-1 is rapidly proteolyzed by DPP-4 and by other enzymes such as endopeptidase 24.11. It is also cleared rapidly by the kidney. As a result, the half-life of GLP-1 is only 1 to 2 minutes. The native peptide, therefore, cannot be used therapeutically. Instead, the approach taken has been to develop metabolically stable analogs or derivatives of GLP-1 that are not subject to the same enzymatic degradation or renal clearance. Five GLP-1 receptor agonists, exenatide, liraglutide, albiglutide, dulaglutide, and lixisenatide are available for clinical use.

a. **Exenatide (exendin 4)** is a GLP-1 receptor agonist isolated from the saliva of the Gila monster (a venomous lizard) that is resistant to DPP-4 action and is cleared by the kidney. Its half life is 2.4 hours and its glucose lowering effect is about 6 hours. Exenatide is dispensed as two fixed-dose pens

(5 and 10 μg). It is injected within the hour before breakfast and before dinner. Patients should be prescribed the 5-μg pen for the first month and then, if tolerated, the dose should be increased to 10 μg twice daily. The drug is not recommended in patients with glomerular filtration rate less than 30 mL/min.

In clinical trials, adding exenatide to the therapeutic regimen of patients with type 2 diabetes who are already taking metformin or a sulfonylurea (or both) further lowered the HbA_{1c} value by 0.4% to 0.6% over a 30-week period. These patients also lost 3 to 6 lb in weight. In an open-label extension study up to 80 weeks, the HbA_{1c} reduction was sustained, and there was further weight loss (to a total loss of ~10 lb). Exenatide LAR is a once weekly preparation that is dispensed as a powder (2 mg). It is suspended in the provided diluent just prior to injection. In comparative clinical trials, the long-acting drug lowers the HbA_{1c} level a little more than the twice daily drug. Low-titer antibodies against exenatide develop in over one-third (38%) of patients, but the clinical effects are not attenuated. High-titer antibodies develop in a subset of patients (~6%), and in about half of these cases, an attenuation of glycemic response has been seen.

b. Liraglutide is a soluble fatty acid acylated GLP-1 analog—with replacement of lysine with arginine at position 34 and the attachment of a C16 acyl chain to a lysine at position 26. The fatty-acyl GLP-1 retains affinity for GLP-1 receptors, but the addition of the C16 acyl chain allows for non-covalent binding to albumin, both hindering DPP-4 access to the molecule and contributing to a prolonged half-life and duration of action. The half-life is approximately 12 hours allowing the drug to be injected once a day. The dosing is initiated at 0.6 mg daily, increased after 1 week to 1.2 mg daily. Some patients may benefit from increasing the dose to 1.8 mg. In clinical trials lasting 26 and 52 weeks, adding liraglutide to the therapeutic regimen (metformin, sulfonyl-urea, thiazolidinedione) of patients with type 2 diabetes further lowered the HbA_{1c} value. Depending on the dose and design of the study, the HbA_{1c} decline was in the range of 0.6% to 1.5%. The patients had sustained weight loss of 1 to 6 lb. Liraglutide at a dose of 3 mg daily has been approved for weight loss.

c. Albiglutide is a human GLP-1 dimer fused to human albumin. It is rendered resistant to DPP-4 action by a glycine substitution for alanine at position 8. The half-life of albiglutide is about 5 days and a steady state is reached after 4-5 weeks of once weekly administration. The usual dose is 30 mg weekly by subcutaneous injection. The dose can be increased to 50 mg weekly, if necessary. The pen contains a lyophilized powder that is reconstituted just prior to administration. Albiglutide monotherapy and combination therapy lowers HbA_{1c} by about 0.8%. Weight loss is much less than with exenatide and liraglutide.

d. Dulaglutide consists of two GLP-1 analog molecules covalently linked to an Fc fragment of human IgG4. The GLP-1 molecule is 90% homologous to native human GLP-1. The amino acid substitutions improve solubility and resist DPP-4 action. The half-life of dulaglutide is about 5 days. The usual dose is 0.75 mg weekly by subcutaneous

injection. The maximum recommended dose is 1.5 mg weekly. Dulaglutide monotherapy and combination therapy lowers HbA_{1c} by about 0.7% to 1.6%. Weight loss ranges from 2 pounds to 7 pounds.

e. Lixisenatide is a modified exendin 5 (deletion of c-terminal proline and addition of 6 lysines) with high affinity for the GLP-1 receptor. Its half-life is about 2 to 4 hours. Dosing is initiated at 10 μg daily and increased to 20 μg daily after two weeks. Lixisenatide monotherapy and combination therapy lowers HbA_{1c} by about 0.7%.

The most frequent adverse reactions of the GLP-1 receptor agonists are nausea (11%-40%), vomiting (4%-13%), and diarrhea (9%-17%). The reactions are more frequent at the higher doses. Albiglutide tends to have the lowest rates of these reactions. In clinical trials about 1% to 5% of participants withdrew from the studies because of the gastrointestinal symptoms.

All the GLP-1 receptor agonists are associated with increased risk of pancreatitis. The FDA reported 30 postmarketing reports of acute pancreatitis in patients taking exenatide. The pancreatitis was severe (hemorrhagic or necrotizing) in 6 instances, and 2 of these patients died. In the liraglutide, albiglutide, and dulaglutide clinical trials, there were 13, 6, and 5 cases of pancreatitis in the drug-treated groups versus 1, 2, and 1 case(s) in the comparator groups, respectively. This translates to about 1.4-2.2 versus 0.6-0.9 cases of pancreatitis per 1000 patient years. Patients taking GLP-1 receptor agonists should be advised to seek immediate medical care if they experience unexplained persistent severe abdominal pain.

The FDA also reported 16 cases of kidney impairment and 62 cases of acute kidney injury taking exenatide. Some of these patients had preexisting kidney disease, and others had one or more risk factors for kidney disease. A number of the patients reported nausea, vomiting, and diarrhea, and it is possible that these side effects caused volume depletion and contributed to the development of renal injury. For this reason, the GLP-1 receptors agonists should be prescribed cautiously in patients with kidney impairment.

GLP-1 receptor agonists stimulate C-cell neoplasia and cause medullary thyroid carcinoma in rats. Human C-cells express very few GLP-1-receptors, and the relevance to human therapy is unclear. The medications, however, should not be used in patients with personal or family history of medullary thyroid carcinoma or multiple endocrine neoplasia (MEN) syndrome type 2.

DPP-4 Inhibitors

An alternative to the use of GLP-1 receptor agonists involves inhibition of the enzyme DPP-4 with prolongation of the action of endogenously released GLP-1 and GIP. Four oral DPP-4 inhibitors, sitagliptin, saxagliptin, linagliptin, and alogliptin are available for the treatment of type 2 diabetes in the United States. An additional DPP-4 inhibitor, vildagliptin, is available in Europe.

a. Sitagliptin when used alone or in combination with other diabetes medicines additionally lowers HbA_{1c} by approximately 0.5%. The usual dose of sitagliptin is 100 mg once daily, but the dose is reduced to 50 mg daily if the calculated creatinine clearance is 30 to 50 mL/min and to 25 mg for clearances less than 30 mL/min. Unlike exenatide, sitagliptin

does not cause nausea or vomiting. It also does not result in weight loss. The main adverse effect appears to be a predisposition to nasopharyngitis or upper respiratory tract infection. A small increase in neutrophil count of approximately 200 cells/mL has also occurred. Since its FDA approval and clinical use, there have been reports of serious allergic reactions to sitagliptin, including anaphylaxis, angioedema, and exfoliative skin conditions, including Stevens-Johnson syndrome. There have also been reports of pancreatitis (88 cases including two cases of hemorrhagic or necrotizing pancreatitis). The frequency of these events is unclear. DPP-4 is a pleiotropic enzyme that inactivates a variety of peptide hormones, neuropeptides, and chemokines and DPP-4 inhibitors have been shown to prolong the action of neuropeptide Y and substance P. Whether its inhibition over a long-term period will have negative consequences is not known.

b. Saxagliptin when added to the therapeutic regimen (metformin, sulfonylurea, thiazolidinedione) of patients with type 2 diabetes further lowered the HbA_{1c} value by about 0.7% to 0.9%. The dose is 2.5 or 5 mg once a day. The 2.5-mg dose should be used in patients with calculated creatinine clearance less than 50 mL/min. The drug is weight neutral. The main adverse reactions were upper respiratory tract infection, naspharyngitis, headache, and urinary tract infection. There is also small reversible dose-dependent reduction in absolute lymphocyte count which remains within normal limits. Hypersensitivity reactions such as urticaria and facial edema occurred in 1.5% of patients on the drug compared to 0.4% with placebo. The metabolism of saxagliptin is through CYP 3A4/5; so strong inhibitors or inducers of CYP 3A4/5 will affect the pharmacokinetics of saxagliptin and its active metabolite. Saxagliptin may increase the risk of heart failure. In a postmarketing study of 16,492 type 2 diabetes patients, there were 289 cases of heart failure in the saxagliptin group (3.5%) and 228 cases in the placebo group (2.8%)—a hazard ratio of 1.27. Patients at the highest risk for failure were those who had a history of heart failure, had elevated levels of plasma N-terminal pro-brain natriuretic peptide (NT-pBNP) or had renal impairment.

c. Alogliptin lowers HbA_{1c} by about 0.5% to 0.6% when added to metformin, sulfonylurea, or pioglitazone. The usual dose is 25 mg orally daily. The 12.5-mg dose is used in patients with calculated creatinine clearance of 30 to 60 mL/min, and 6.25 mg for clearance less than 30 mL/min. In clinical trials, pancreatitis occurred in 11 of 5902 patients on alogliptin (0.2%) and in 5 of 5183 patients receiving all comparators (<0.1%). There have been reports of hypersensitivity reactions (anaphylaxis, angioedema, Stevens-Johnson syndrome). Cases of hepatic failure have been reported but it is unclear if alogliptin was the cause. The medication, however, should be discontinued in the event of liver failure.

In a large postmarketing study, alogliptin, like saxagliptin, was associated with a slightly increased rate of heart failure.

d. Linagliptin lowers HbA_{1c} by about 0.4% to 0.6% when added to metformin, sulfonylurea, or pioglitazone. The dose is 5 mg orally daily, and since it is primarily excreted unmetabolized through the bile, no dose adjustment is needed in patients with renal failure. The adverse reactions include

nasopharyngitis and hypersensitivity reactions (urticaria, angioedema, localized skin exfoliation, bronchial hyperreactivity). In one study, there were eight cases of pancreatitis in 4687 patients exposed to drug (4311 patient years) with 0 cases in 1183 patients receiving placebo (433 patient years).

e. Vildagliptin, like the other DPP-4 inhibitors, lowers HbA_{1c} by about 0.5% to 1% when added to the therapeutic regimen of patients with type 2 diabetes. The dose is 50 mg once or twice a day. The adverse reactions include upper respiratory tract infections, nasopharyngitis, dizziness, and headache. Rare cases of hepatic dysfunction including hepatitis have been reported. Liver function testing is recommended, quarterly the first year, and periodically thereafter.

Animal studies using much higher doses of DPP-4 inhibitors and GLP-1-receptor agonists than are used in humans caused expansion of pancreatic ductal glands and generation of premalignant pancreatic intraepithelial (PanIN) lesions that have the potential to progress to pancreatic adenocarcinoma. There is, however, currently no evidence that these drugs cause pancreatic cancer in humans.

5. SODIUM-GLUCOSE COTRANSPORTER 2 INHIBITORS

Glucose is freely filtered by the renal glomeruli and is reabsorbed in the proximal tubules by the action of sodium-glucose cotransporters (SGLT). Sodium-glucose cotransporter 2 (SGLT2) accounts for about 90% of glucose reabsorption and its inhibition causes glycosuria in people with diabetes, lowering plasma glucose levels. The SGLT2 inhibitors canagliflozin, dapagliflozin, and empagliflozin are approved for clinical use in the United States.

a. Canagliflozin reduces the threshold for glycosuria from a plasma glucose threshold of ~180 mg/dL to 70–90 mg/dL. It has been shown to reduce HbA_{1c} by 0.6% to 1% when used alone or in combination with other oral agents or insulin. It also results in modest weight loss of 2–5 kg. The usual dose is 100 mg daily but up to 300 mg daily can be used in patients with normal kidney function.

b. Dapagliflozin is an SGLT2 inhibitor that has been shown to reduce HbA_{1c} levels by 0.5% to 0.8% when used alone or in combination with other oral agents or insulin. It also results in modest weight loss of about 2 to 4 kg. The usual dose is 10 mg daily but 5 mg daily is the recommended initial dose in patients with hepatic failure.

c. Empagliflozin reduces HbA_{1c} by 0.5% to 0.7% when used alone or in combination with other oral agents or insulin. It also results in modest weight loss of about 2 to 3 kg. The usual dosage is 10 mg daily but a higher dose of 25 mg daily can be used. In a postmarketing multinational study of 7020 type 2 patients with known cardiovascular disease, the addition of empagliflozin was associated with a lower primary composite outcome of death from cardiovascular causes, nonfatal myocardial infarction, or non-fatal stroke (hazard ratio 0.86, $p = 0.04$). The mechanism(s) regarding the benefit remains unclear. Weight loss, lower blood pressure, and diuresis may have played a role since there were fewer deaths from heart failure in the treated group whereas the rates of myocardial infarction were unaltered. Additional studies with empagliflozin and other SGLT2 inhibitors are needed to confirm this result.

As might be expected, the efficacy of the SGLT2 inhibitors is reduced in chronic kidney disease. They can also increase creatinine and decrease eGFR, especially in patients with kidney impairment. Canagliflozin is contraindicated in patients with eGFR less than 45 mL/min/1.73 m^2. The main side effects are increased incidence of genital mycotic infections and urinary tract infections affecting ~8% to 9% of patients. There have also been reports of cases of pyelonephritis and septicemia requiring hospitalization. The glycosuria can cause intravascular volume contraction and hypotension.

Canagliflozin has been reported to cause a decrease in bone mineral density at the lumbar spine and the hip. An increase in fractures by about 30% was observed in patients on canagliflozin in a pooled analysis of 8 clinical trial (mean duration 68 weeks). It is likely that the effect on the bones is a class effect and not restricted to canagliflozin. A modest increase in upper limb fractures was observed with canaglifloxin therapy. It is not known if this is due to an effect on bone strength or related to falls due to hypotension. All the SGLT2 inhibitors cause a modest increase in LDL cholesterol levels (3%-8%). Also, in clinical trials, patients taking dapagliflozin had higher rates of breast cancer (nine cases vs none in comparator arms) and bladder cancer (10 cases vs 1 in placebo arm). These cancer rates exceeded the expected rates in age-matched reference diabetes population.

Cases of diabetic ketoacidosis have been reported with off label use of SGLT2 inhibitors in patients with type 1 diabetes. Type 1 patients are taught to give less insulin if their glucose levels are not elevated. SGLT2 inhibitors lower glucose levels by changing the renal threshold for glucose reabsorption (ie, increased glucosuria) and not through insulin action. Type 1 patients on an SGLT2 inhibitor, because the glucose levels are not elevated, may either withhold or reduce their insulin doses to such a degree as to induce ketoacidoisis, SGLT2 inhibitors should not be used in patients with type 1 diabetes and in those patients labeled as having type 2 diabetes but who are very insulin deficient and ketosis-prone.

6. OTHERS

a. Pramlintide is a synthetic analog of islet amyloid polypeptide (IAPP) that when given subcutaneously (1) delays gastric emptying, (2) suppresses glucagon secretion, and (3) decreases appetite. It is approved for use both in type 1 and insulin-treated type 2 patients. In 6-month clinical studies with type 1 and insulin-treated type 2 patients, those on the drug had approximately a 0.4% reduction in HbA$_{1c}$ and lost about 1.7 kg compared to placebo. The HbA$_{1c}$ reduction was sustained for 2 years, but some of the weight was regained. The drug is given immediately before the meal by injection. Hypoglycemia is the most concerning adverse event, and it is recommended that the short-acting or premixed insulin doses be reduced by 50% when the drug is started. Since the drug slows gastric emptying, recovery from hypoglycemia can be a problem because of delay in absorption of the fast-acting carbohydrate. Nausea was the other main side effect, affecting 30% to 50% of subjects. It tended to improve with time. In patients with type 1 diabetes, pramlintide is initiated at the dose of 15 μg before each meal and titrated by 15 μg increments to a maintenance dose of 30 or 60 μg before each meal. In patients with type 2 diabetes, the initiation dose is 60 μg premeals increased to 120 μg in 3 to 7 days if no significant nausea occurs.

b. Bromocriptine, a dopamine 2 receptor agonist, has been shown to modestly lower HbA$_{1c}$ by 0.1% to 0.5% when compared to baseline and 0.4% to 0.5% compared to placebo. The tablet dose is 0.8 mg and the daily dose is 2 (1.6 g) to 6 (4.8 mg) tablets as tolerated. Common side effects are nausea, vomiting, dizziness, and headache.

c. Colesevelam, the bile acid sequestrant, when added to metformin or sulfonylurea or insulin lowered HbA$_{1c}$ 0.3% to 0.4% when compared to baseline and 0.5% to 0.6% compared to placebo. HbA$_{1c}$ lowering, however, was not observed in a single monotherapy clinical trial comparing colesevelam to placebo. Colesevelam use is associated with ~20% increase in triglyceride levels. Other adverse effects include constipation and dyspepsia.

With their modest glucose lowering and significant side effects, using bromocriptine or colesevelam to treat diabetes is not recommended.

Drug Combinations

Several drug combinations are available in different dose sizes including glyburide-metformin (Glucovance), glipizide-metformin (Metaglip), repaglinide-metformin (Prandi-Met), rosiglitazone-metformin (Avandamet), pioglitazone-metformin (ACTOplus Met), rosiglitazone and glimepiride (Avandaryl), pioglitazone and glimepiride (Duetact), sitagliptin-metformin (Janumet), saxagliptin-metformin extended release (kombiglyze XR), linagliptin and metformin (Jentadueto), alogliptin and metformin (Kazano), alogliptin and pioglitazone (Oseni), dapagliflozin and metformin (Xigduo), canagliflozin and metformin (Invokamet), empagliflozin and metformin (Synjardy), empagliflozin and linagliptin (Glyxambi), and lixisenatide and insulin glargine (Soliqua). These combinations, however, limit the clinician's ability to optimally adjust dosage of the individual drugs and for that reason are not recommended.

7. INSULIN

Insulin is indicated for individuals with type 1 diabetes as well as for those with type 2 diabetes whose hyperglycemia does not respond to diet therapy and other diabetes drugs.

Human insulin is dispensed as either regular (R) or Neutral Protamine Hagedorn (NPH) formulations. Animal insulins are no longer available in the United States. Six analogs of human insulin—three rapidly acting (insulin lispro, insulin aspart, and insulin glulisine) and three long-acting (insulin glargine, insulin detemir, and insulin degludec) are available for clinical use. All currently available insulins contain less than 10 ppm of proinsulin and are labeled as "purified." These purified insulins preserve their potency, so that refrigeration is recommended but not crucial. During travel, reserve supplies of insulin can be readily transported for weeks without losing potency if protected from extremes of heat or cold. All the insulins in the United States are available in a concentration of 100 U/mL (U100) and dispensed in 10 mL vials or 0.3 mL cartridges or prefilled disposable pens. Several insulins are now available at higher concentrations—insulin glargine 300 U/mL (U300), insulin degludec 200 U/mL (U200),
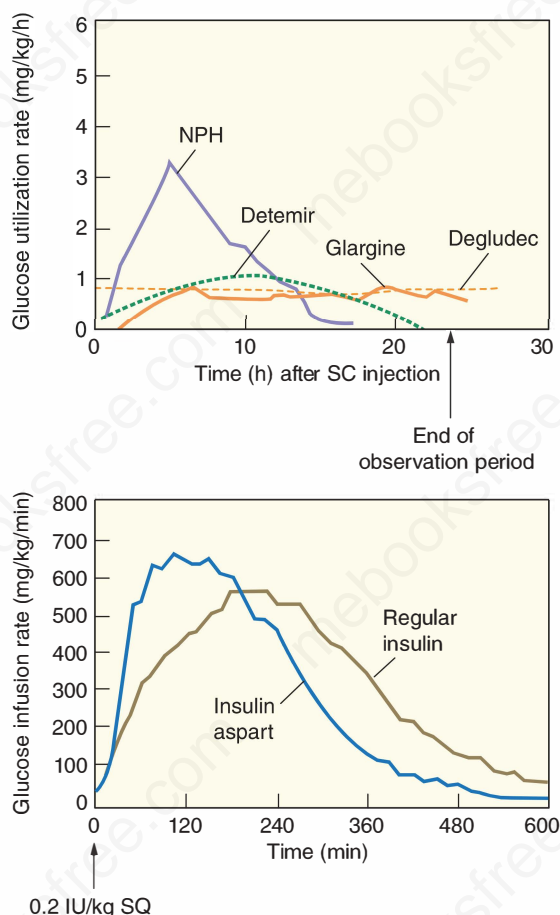
FIGURE 17–12 Extent and duration of action of various types of insulin (in a fasting diabetic). Duration of action may be extended considerably when the dose of a given insulin formulation increases above the average therapeutic doses depicted here.

insulin lispro 200 U/mL (U200), and regular insulin 500 U/mL (U500). (Figure 17–12; Tables 17–13 and 17–14).

It is important to recognize that values given for time of onset of action, peak effect, and duration of action are only approximate and that there is great variability in these parameters from patient to patient and even in a given patient depending on the size of the dose, the site of injection, the degree of exercise, the avidity of circulating anti-insulin antibodies, and other less well-defined variables.

Short-Acting Insulin Preparations

The short-acting preparations are regular insulin and the rapidly acting insulin analogs. They are dispensed as clear solutions at neutral pH and contain small amounts of zinc to improve their stability and shelf life.

A. Regular insulin   Regular insulin is a short-acting, soluble crystalline zinc insulin whose hypoglycemic effect appears within 30 minutes after subcutaneous injection, peaks at about 2 hours, and lasts for about 5 to 7 hours when usual quantities (ie, 5-15 U) are administered. Intravenous infusions of regular

TABLE 17–13 Summary of bioavailability characteristics of the insulins.

Insulin Preparations	Onset of Action	Peak Action	Effective Duration
Insulins lispro, aspart, glulisine	5-15 min	1-1.5 h	3-4 h
Human regular	30-60 min	2 h	6-8 h
Technosphere inhaled insulin	5-15 min	1 h	3 h
Human NPH	2-4 h	6-7 h	10-20 h
Insulin glargine	0.5-1 h	Flat	~24 h
Insulin detemir	0.5-1 h	Flat	17 h
Insulin degludec	0.5-1.5 h	Flat	>42 h

insulin are particularly useful in the treatment of diabetic keto-acidosis and during the perioperative management of insulin-requiring diabetics.

For very insulin-resistant subjects who would otherwise require large volumes of insulin solution, a U500 preparation of human regular insulin is available both in a vial form and a disposable pen. If the vial form is used, a U500 insulin syringe is available to measure doses. The disposable pen dispenses the regular insulin in 5 unit increments.

B. Rapidly acting insulin analogs   *Insulin lispro* (Humalog) is an insulin analog where the proline at position B28 is reversed with the lysine at B29. *Insulin aspart* (Novolog) is a single

TABLE 17–14 Insulin preparations available in the United States.[a]

Short Acting Insulin Preparations
Insulin lispro (Humalog, Lilly)
Insulin aspart (Novolog, Novo Nordisk)
Insulin glusisine (Apidra, Sanofi Aventis)
Regular insulin (Lilly, Novo Nordisk)
Technosphere inhaled regular insulin (Afrezza, Mannkind)

Long acting Insulin Preparations
NPH insulin (Lilly, Novo Nordisk)
Insulin glargine (Lantus, Sanofi Aventis)
Insulin detemir (Levemir, Novo Nordisk)
Insulin degludec (Tresiba, Novo Nordisk)

Premixed Insulins
   70% NPH/30% regular (70/30 insulin, Lilly, Novo Nordisk)
   75% insulin lispro/protamine (NPL) 25% insulin lispro
      (Humalog Mix 75/25, Lilly)
   50% NPL/50% insulin lispro (Humalog Mix 50/50, Lilly)
   70% insulin aspart/protamine/30% insulin aspart
      (Novolog Mix 70/30, Novo Nordisk)
   70% insulin degludec/30% insulin aspart
      (Ryzodeg, Novo Nordisk)

[a]All types of insulin available in the United States are recombinant human or human insulin analog origin. All the insulins are dispensed at U100 concentration. There is an additional U500 preparation of regular insulin; U300 preparation of insulin glargine; U200 preparations of insulin lispro and insulin degludec.

substitution of proline by aspartic acid at position B28. In *Insulin glulisine* (Apidra) the asparagine at position B3 is replaced by lysine and the lysine in position B29 by glutamic acid. These three analogs have less tendency to form hexamers compared to human insulin. When injected subcutaneously, the analogs quickly dissociate into monomers and are absorbed very rapidly, reaching peak serum values in as soon as 1 hour—in contrast to regular human insulin, whose hexamers require considerably more time to dissociate and become absorbed. The amino acid changes in these analogs do not interfere with their binding to the insulin receptor, with the circulating half-life, or with their immunogenicity, which are all identical to those of human regular insulin.

Clinical trials have demonstrated that the optimal times of preprandial subcutaneous injection of comparable doses of the rapid-acting insulin analogs and of regular human insulin are 20 minutes and 60 minutes before the meal, respectively. Although this more rapid onset of action has been welcomed as a great convenience by patients with diabetes who object to waiting as long as 60 minutes after injecting regular human insulin before they can begin their meal, patients must be taught to ingest adequate absorbable carbohydrate early in the meal to avoid hypoglycemia during the meal. Another desirable feature of rapidly acting insulin analogs is that their duration of action remains at about 4 hours irrespective of dosage. This contrasts with regular insulin, whose duration of action is prolonged when larger doses are used.

The rapidly acting analogs are also commonly used in pumps. In a double-blind crossover study comparing insulin lispro with regular insulin in insulin pumps, persons using insulin lispro had lower HbA_{1c} values and improved postprandial glucose control with the same frequency of hypoglycemia. In the event of pump failure, however, users of the rapidly acting insulin analogs will have more rapid onset of hyperglycemia and ketosis.

While insulin aspart has been approved for intravenous use (eg, in hyperglycemic emergencies), there is no advantage in using insulin aspart over regular insulin by this route. A U200 concentration of insulin lispro is available in a disposable prefilled pen. The only advantage of the U200 over the U100 insulin lispro preparation is that it delivers the same dose in half the volume.

Long-Acting Insulin Preparations

a. Neutral Protamine Hagedorn (NPH), or isophane, insulin is an intermediate-acting insulin in which the onset of action is delayed by combining two parts of soluble crystalline zinc insulin with one part protamine zinc insulin. The mixture has equivalent concentrations of protamine and insulin, so that neither is in excess (isophane). Its onset of action is delayed by 2 to 4 hours, and its peak response is generally reached in about 8 to 10 hours. Because its duration of action is often less than 24 hours (with a range of 10-20 hours), most patients require at least two injections daily to maintain a sustained insulin effect.

Flocculation of suspended particles may occasionally frost the sides of a bottle of NPH insulin or clump within bottles from which multiple small doses are withdrawn over a prolonged period. This instability is a rare phenomenon and might occur less frequently if NPH human insulin were refrigerated when not in use and if bottles were discarded after 1 month of use. Patients should be vigilant for early signs of frosting or clumping of the NPH insulin, because it indicates a pronounced loss of potency. Several cases of diabetic ketoacidosis have been reported in patients with type 1 diabetes who had been inadvertently injecting this denatured insulin.

b. Insulin glargine is an insulin analog in which the asparagine at position 21 of the A chain of the human insulin molecule is replaced by glycine and two arginines are added to the carboxyl terminal of the B chain. The arginines raise the isoelectric point of the molecule close to neutral, making it more soluble in an acidic environment. In contrast, human insulin has an isoelectric point of pH 5.4. Insulin glargine is a clear insulin that, when injected into the neutral pH environment of the subcutaneous tissue, forms microprecipitates that slowly release the insulin into the circulation. It lasts for about 24 hours without any pronounced peaks and is given once a day to provide basal coverage. This insulin cannot be mixed with the other insulins because of its acidic pH. When this insulin was given as a single injection at bedtime to type 1 diabetes patients, fasting hyperglycemia was better controlled when compared with bedtime NPH insulin. The clinical trials also suggest that there may be less nocturnal hypoglycemia with this insulin when compared with NPH insulin.

In one clinical trial involving patients with type 2 diabetes, insulin glargine was associated with a slightly more rapid progression of retinopathy when compared with NPH insulin. The frequency was 7.5% with the analog and 2.7% with the NPH. This observation, however, was not confirmed in a 5-year open label prospective study of 1024 subjects randomized to NPH or insulin glargine. In *in vitro* studies, insulin glargine has a sixfold greater affinity for the IGF-1 receptor compared with the human insulin. There has also been a report that insulin glargine has increased mitogenicity compared with human insulin in a human osteosarcoma cell line. Circulating levels of insulin glargine, however, are low, and the clinical significance of these observations is not yet clear. An observational study from Germany of 127,031 patients who had exposure to regular insulin, short-acting insulin analogs, and insulin glargine reported a strong correlation between increased insulin dose and cancer risk. Moreover insulin glargine, dose for dose, appeared to carry a higher risk than regular insulin. Subsequent epidemiologic studies, however, failed to confirm this observation. A more concentrated form of insulin glargine (U300) is available as an insulin pen. In pharmacodynamic studies in type 1 patients, the U300 compared to the U100 preparation had approximately 5 hours longer duration of action. In clinical trials in type 1 patients, U300 use did not result in better control or reduce the rates of hypoglycemia. Although limited clinical data suggest that insulin glargine is safe in pregnancy, it is not approved for this use.

c. Insulin detemir is an insulin analog in which the threonine at position 30 of the B chain has been removed and a 14-C fatty acid chain (tetradecanoic acid) is attached to the lysine at position 29 by acylation. The fatty acid chain makes the molecule more lipophilic than native insulin, and the addition of zinc stabilizes the molecule and leads to formation of hexamers. After injection, self-association at the injection site and albumin binding in the circulation via the fatty acid side chain lead to slower distribution to peripheral target

tissues and prolonged duration of action. The affinity of insulin detemir is four- to fivefold lower than that of human soluble insulin and, therefore, the U100 formulation of insulin detemir has a concentration of 2400 nmol/mL compared to 600 nmol/mL for NPH. The duration of action for insulin detemir is about 17 hours at therapeutically relevant doses. It is recommended that the insulin is injected once or twice a day to achieve a stable basal coverage. Apparently this insulin has been reported to have lower within-subject pharmacodynamic variability compared to NPH insulin and insulin glargine. *In vitro* studies do not suggest any clinically relevant albumin binding interactions between insulin detemir and fatty acids or protein-bound drugs. Because there is a vast excess (~400,000) of albumin-binding sites available in plasma per insulin detemir molecule, it is unlikely that hypoalbuminemic disease states will affect the ratio of bound to free insulin detemir.

d. Insulin degludec In this insulin analog, the threonine at position B30 has been removed and the lysine at position B29 is conjugated to hexadecanoic acid via a gamma-L-glutamyl spacer. In the vial, in the presence of phenol and zinc, the insulin is in the form of dihexamers but when injected subcutaneously, it self-associates into large multi-hexameric chains consisting of thousands of dihexamers. The chains slowly dissolve in the subcutaneous tissue and insulin monomers are steadily released into the systemic circulation. The half-life of the insulin is 25 hours. Its onset of action is in 30 to 90 minutes and its duration of action is more than 42 hours. It is recommended that the insulin be injected once or twice a day to achieve a stable basal coverage. Insulin degludec is available in two concentrations, U100 and U200 and dispensed in prefilled disposable pens.

Insulin Mixtures

Patients with type 2 diabetes can sometimes achieve reasonable glucose control with just prebreakfast and presupper injections of mixtures of short acting and NPH insulins. The regular insulin or rapidly acting insulin analog is withdrawn first, then the NPH insulin and then injected immediately. Stable premixed insulins (70% NPH and 30% regular) are available as a convenience to patients who have difficulty mixing insulin because of visual problems or insufficient manual dexterity.

With increasing use of rapid-acting insulin analogs as a preprandial insulin, it has become evident that combination with an intermediate-acting or long-acting insulin is essential to maintain postabsorptive glycemic control. It has been demonstrated that insulin lispro can be acutely mixed with NPH insulin without affecting its rapid absorption. Premixed preparations of insulin lispro and NPH insulin however are unstable because of exchange of insulin lispro with the human insulin in the protamine complex. Consequently, over time, the soluble component becomes a mixture of regular and insulin lispro at varying ratios. In an attempt to remedy this, intermediate insulin composed of isophane complexes of protamine with insulin lispro was developed and given the name NPL (neutral protamine lispro). Premixed combinations of NPL and insulin lispro are now available for clinical use (Humalog Mix 75/25 and Humalog Mix 50/50). These mixtures have a more rapid onset of glucose-lowering activity compared with 70% NPH/30% regular human

insulin mixture and can be given within 15 minutes before or after starting a meal. Similarly, a 70% insulin aspart protamine/30% insulin aspart (NovoLogMix 70/30) is now available. The main advantages of these new mixtures are that (1) they can be given within 15 minutes of starting a meal and (2) they are superior in controlling the postprandial glucose rise after a carbohydrate-rich meal. These benefits have not translated into improvements in HbA$_{1c}$ levels when compared with the usual 70% NPH/30% regular mixture.

The longer-acting insulin analogs, insulin glargine or insulin detemir, cannot be acutely mixed with either regular insulin or the rapid-acting insulin analogs. Insulin degludec, however, can be mixed and is available as 70% insulin degludec/30% insulin aspart and is injected once or twice a day.

Methods of Insulin Administration

a. Insulin syringes and needles—Disposable plastic syringes with needles attached are available in 1-mL (100 units), 0.5-mL (50 units), and 0.3-mL (30 units) sizes. The "low-dose" 0.3 mL syringes are popular because many patients with diabetes do not take more than 30 units of insulin in a single injection except in rare instances of extreme insulin resistance. Three lengths of needles are available: 6 mm, 8 mm, and 12.7 mm. Long needles are preferable in obese patients to reduce variability of insulin absorption. The needles are of 28, 30, and 31 gauges. The 31 gauge needles are almost painless. Many patients reuse the syringes up to 3 to 5 times, maintaining adequate sterility by recapping between uses.

Any part of the body covered by loose skin can be used as an injection site, including the abdomen, thighs, upper arms, flanks, and upper-outer quadrants of the buttocks. Cleaning with alcohol is not necessary as long as the skin is clean. Rotation of sites is recommended to avoid delayed absorption when lipohypertrophy occurs from repeated use of a single site. In general, regular insulin is absorbed more rapidly from upper regions of the body such as the deltoid area or the abdomen rather than from the thighs or buttocks. Exercise appears to facilitate insulin absorption when the injection site is adjacent to the exercising muscle. For most patients, the abdomen is the recommended region for injection because it provides a considerable area in which to rotate sites, and there may be less variability of absorption with exercise than when the thigh or deltoid areas are used. The effect of anatomic regions appears to be much less pronounced with the analogs.

b. Insulin pen injector devices—Insulin pens eliminate the need for carrying insulin vials and syringes. Cartridges of insulin lispro, insulin aspart, and insulin glargine are available for reusable pens (Lilly, Novo Nordisk, and Owen Mumford). Disposable prefilled pens are also available for regular insulin (U100 and U500), insulin lispro, insulin aspart, insulin glulisine, insulin detemir, insulin glargine, insulin degludec, NPH, 70% NPH/30% regular, 75% NPL/25% insulin lispro, 50% NPL/50% insulin lispro, 70% insulin aspart protamine/30% insulin aspart, and 70% insulin degludec/30% insulin aspart. Pen needles are available in 29, 31, and 32 gauges and 4, 5, 6, 8, and 12.7 mm lengths.

c. Insulin pumps—In the United States, Medtronic MiniMed, Insulet, Animas, Roche, and Tandem make battery

operated subcutaneous insulin infusion pumps. These pumps are small (about the size of a pager) and easy to program. They have many features, including the ability to record a number of different basal rates throughout a 24-hour period and adjust the time over which bolus doses are given. They are able also to detect pressure build-up if the catheter is kinked. The catheter connecting the insulin reservoir to the subcutaneous cannula can be disconnected so the patient can remove the pump temporarily (eg, for bathing). Ominpod (Insulet Corporation) is an insulin infusion system in which the insulin reservoir and infusion set are integrated into one unit (pod), so there is no catheter (electronic patch pump). The pod, placed on the skin, delivers subcutaneous basal and bolus insulin based on wirelessly transmitted instructions from a personal digital assistant.

The great advantage of **continuous subcutaneous insulin infusion (CSII)** is that it allows for establishment of a basal profile tailored to the patient allowing for better overnight and between meals glucose control. The ability to adjust the basals makes it easier for the patient to manage glycemic excursions that occur with exercise. The pumps have software that can assist the patient to calculate boluses based on glucose reading and carbohydrates to be consumed. They also keep track of the time elapsed since last insulin bolus and the patient is reminded of this when he or she attempts to give additional correction bolus before the effect of the previous bolus has worn off (*insulin on board* feature). This feature reduces the risk of overcorrecting and subsequent hypoglycemia.

CSII therapy is appropriate for patients who are motivated, mechanically adept, educated about diabetes (diet, insulin action, treatment of hypo- and hyperglycemia), and willing to monitor their blood glucose four to six times a day. Known complications of CSII include ketoacidosis, which can occur when insulin delivery is interrupted, and skin infections. Another major disadvantage is the cost and the time demanded of physicians and staff in initiating therapy. Almost all patients use the rapid-acting insulin analogs in their pumps. Clinical trials have shown that when compared with regular insulin, subjects using rapid-acting insulin analogs in pumps had lower HbA$_{1c}$ values and improved postprandial glucose control with the same frequency of hypoglycemia.

V-go (Valeritas) is a mechanical patch pump designed specifically for people with type 2 diabetes who employ a basal/bolus insulin regimen. The device is preset to deliver one of three fixed and flat basal rates (20, 30, or 40 units) for 24 hours (at which point it must be replaced) and there is a button that delivers two units per press to help cover meals.

d. **Inhaled insulin**—Technosphere insulin (Afrezza) is a dry-powder formulation of regular insulin that can be inhaled. It consists of 2 to 2.5 μm crystals of the excipient, fumaryl diketopiperazine, that provide large surface area for adsorption of proteins like insulin. Pharmacokinetic studies show that technosphere insulin is rapidly absorbed with peak insulin levels reached in 12 to 15 minutes and declining to baseline in 3 hours. Pharmacodynamic studies show that median time to maximum effect with inhaled insulin is approximately 1 hour and declines to baseline by about 3 hours. In contrast, the median time to maximum effect with subcutaneous insulin lispro is about 2 hours and declines to baseline by 4 hours. In clinical trials, technosphere insulin

combined with basal insulin was as effective in glucose lowering as rapid-acting insulin analogs combined with basal insulin. It is formulated as a single use color coded cartridge delivering 4, 8, or 12 units immediately before the meal. The manufacturer provides a dose conversion table; patients injecting up to 4 units of rapid-acting insulin analog should use the 4 unit cartridge. Those injecting 5 to 8 units should use the 8-unit cartridge. If the dose is 9 to 12 units of rapid-acting insulin premeal then one 4-unit cartridge and one 8-unit cartridge or one 12-unit cartridge should be used. The inhaler is about the size of a referee's whistle.

The most common adverse reaction of the inhaled insulin was a cough affecting about 27% of patients. A small decrease in pulmonary function (forced expiratory volume in 1 second [FEV1]) was seen in the first 3 months of use, which persisted over 2 years of follow-up. Inhaled insulin is contraindicated in smokers and patients with chronic lung disease, such as asthma and chronic obstructive pulmonary disease. Spirometry should be performed to identify potential lung disease prior to initiating therapy. During the clinical trials, there were two cases of lung cancer in patients who were taking inhaled insulin and none in the comparator-treated patients. All the patients in whom lung cancer developed had a history of prior cigarette smoking. There were also two cases of squamous cell carcinoma of the lung in nonsmokers exposed to inhaled insulin; these cases occurred after completion of the clinical trials. Cases of lung cancer were also reported in cigarette smokers using a previously available inhaled insulin preparation (Exubera). The incidence rate in the Exubera treated group was 0.13 per 1000 patient years and 0.03 per 1000 patient years in the comparator-treated group.

e. **Pancreas transplantation**—Type 1 patients who are candidates for a kidney transplant should be evaluated for a pancreas transplant. Contraindications include age greater than 55 years old, uncorrectable coronary artery disease; extensive peripheral vacular disease, and significant obesity (weight >100 kg). The pancreas transplant may occur simultaneously (simultaneous pancreas/kidney) or after kidney transplant (pancreas after kidney).

Patients undergoing simultaneous pancreas and kidney transplantation have an 85% chance of pancreatic graft survival at 1 year and 69% at 5 years. Solitary pancreatic transplantation in the absence of a need for renal transplantation should be considered only in those rare patients who fail all other insulin therapeutic approaches and who have frequent severe hypoglycemia or who have life-threatening complications related to their lack of metabolic control. Pancreas transplant alone graft survival is 78% at 1 year and 54% at 5 years.

f. **Islet cell transplantation**—Total pancreatectomy is curative for severe pain syndrome associated with chronic pancreatitis. The pancreatectomy, however, results in surgical diabetes. Harvesting islets from the removed pancreas and autotransplanting them into the liver (via portal vein) can prevent the development of diabetes or result in "mild" diabetes (partial islet function) that is easier to manage. Since the islets are autologous, no immunosuppression is required. The number of islets transplanted is the main predictor of insulin independence.

People with type 1 diabetes can become insulin independent after receiving an islet cell transplant. The islets are isolated from a donor pancreas (alloislet transplant). The

islets are infused into the portal vein using a percutaneous transhepatic approach, and they lodge in the liver releasing insulin in response to physiologic stimuli. Long-term immunosuppression is necessary to prevent allograft rejection and to suppress the autoimmune process that led to the disease in the first place. Insulin independence for more than 5 years has been demonstrated in patients who got anti-CD3 antibody or anti-thymocyte globulin for induction of immunosuppression and calcineurin inhibitors, mTor inhibitors, and mycophenolate mofetil as maintenance immunosuppression. One major limitation is the need for more than one islet infusion to achieve insulin independence. This is because of significant loss of islets during isolation and the period prior to engraftment. Alloislet transplantation is currently an experimental procedure and widespread application will depend on improving insulin independence rates with one infusion and also demonstrating that the long-term outcomes are as good as those of pancreas transplant alone.

STEPS IN THE MANAGEMENT OF THE DIABETIC PATIENT

History and Examination

A complete history is taken and physical examination is performed for diagnostic purposes and to rule out the presence of coexisting or complicating disease. Nutritional status should be noted, particularly if catabolic features such as progressive weight loss are present despite a normal or increased food intake. The family history should include not only the incidence but also the age at onset of diabetes in other members of the family, and it should be noted whether affected family members were obese, whether they required insulin, and whether they developed complications from their diabetes. Other factors that increase cardiovascular risk, such as a smoking history, presence of hypertension or hyperlipidemia, or oral contraceptive pill use should be documented.

A careful physical examination should include baseline height and weight, pulse rate, and blood pressure. If obesity is present, it should be characterized as to its distribution, and a waist-to-hip ratio should be recorded. All peripheral arterial pulses should be examined, noting whether bruits or other signs of atherosclerotic disease are present. Neurologic and ophthalmologic examinations should be performed, with emphasis on investigation of abnormalities that may be related to diabetes, such as neovascularization of the retina or stocking/glove sensory loss in the extremities.

Laboratory Diagnosis

Laboratory diagnosis should include documentation of the presence of fasting hyperglycemia (plasma glucose ≥126 mg/dL [7 mmol/L]), postprandial (post-glucose tolerance test) values consistently greater than or equal to 200 mg/dL (11.1 mmol/L) or HbA_{1c} greater than or equal to 6.5%. An attempt should be made to characterize the diabetes as type 1 or type 2, based on the clinical features present and on whether ketonuria accompanies the glycosuria. For the occasional patient, measurement of ICA 512, GAD, insulin and zinc transporter 8 antibodies can help in distinguishing between type 1 and type 2 diabetes. Many newly diagnosed patients with type 1 diabetes still have significant endogenous insulin production, and C-peptide levels may not reliably distinguish between type 1 and type 2 diabetes. Other baseline laboratory measurements that should be made part of the record include hemoglobin A_{1c}, lipid profile, serum creatinine and electrolytes, complete blood count, electrocardiogram, and urine albumin measurement (type 2 patient).

Patient Education and Self-Management Training

Since diabetes is a lifelong disorder, education of the patient and the family is probably the most important obligation of the clinician who provides initial care. The best persons to manage a disease that is affected so markedly by daily fluctuations in environmental stress, exercise, diet, and infections are the patients themselves and their families. It must be remembered that education is necessary not only for patients with newly diagnosed diabetes and their families, but also for patients with diabetes of any duration who may never have been properly educated about their disorder or who may not be aware of advances in diabetes management. The *teaching curriculum* should include explanations by the physician or nurse of the nature of diabetes and its potential acute and chronic hazards and how they can be recognized early and prevented or treated. Self-monitoring of blood glucose should be emphasized, especially in insulin-requiring diabetic patients, and instructions must be given on proper testing and recording of data. Patients must also be helped to accept the fact that they have diabetes; until this difficult adjustment is made, efforts to cope with the disorder are likely to be futile. Counseling should be directed at avoidance of extremes such as compulsive rigidity or self-destructive neglect.

Patients on insulin should have an understanding of the actions of basal and bolus insulins. They should be taught how to determine whether the basal dose is appropriate and how to adjust the rapidly acting insulin dose for the carbohydrate content of a meal. Patients and their families or friends should also be taught to recognize signs and symptoms of hypoglycemia and how to institute appropriate therapy for hypoglycemic reactions. Strenuous exercise can precipitate hypoglycemia, and patients should know how much to reduce their insulin dosage in anticipation of strenuous activity or to take supplemental carbohydrate. Injection of insulin into a site farthest away from the muscles most involved in exercise may help ameliorate exercise-induced hypoglycemia, since insulin injected in the proximity of exercising muscle may be more rapidly mobilized. Exercise training also increases the effectiveness of insulin and insulin doses should be adjusted accordingly. Because infections, particularly pyogenic ones with fever and toxemia, provoke a marked increase in insulin requirements, patients must be taught how to appropriately administer supplemental rapid-acting insulin as needed to correct hyperglycemia during infections.

Type 2 diabetes on noninsulin medications should be informed about the time of onset, peak action, duration of action, and any

adverse effects of pharmacologic agents being used. They should also learn to inquire about possible drug interactions whenever any new medications are added to their regimens.

The targets for blood glucose control should be elevated appropriately in elderly patients since they have the greatest risk if subjected to hypoglycemia and the least long-term benefit from more rigid glycemic control. Patients should be provided advice on personal hygiene, including detailed instructions on foot and dental care. All infections (especially pyogenic ones) provoke the release of high levels of insulin antagonists such as catecholamines or glucagon and thus bring about a marked increase in insulin requirements. Patients who use oral agents may decompensate and temporarily require insulin. All patients receiving therapy that can cause hypoglycemia should wear a MedicAlert bracelet or necklace that clearly states that insulin or an oral sulfonylurea drug is being taken. Patients should be told about community agencies, such as American Diabetes Association chapters, that can serve as a continuing source of instruction.

Finally, vigorous efforts should be made to persuade new diabetics who smoke to give up the habit, since large vessel peripheral vascular disease and debilitating retinopathy are less common in nonsmoking diabetic patients.

Specific Therapy

A reasonable aim of therapy is to approach normal glycemic excursions without provoking severe or frequent hypoglycemia. Criteria for "acceptable" control includes the following: (1) blood glucose levels of 90 to 130 mg/dL (5-7.2 mmol/L) before meals and after an overnight fast, (2) levels no higher than 180 mg/dL (10 mmol/L) 1 hour after meals and 150 mg/dL (8.3 mmol/L) 2 hours after meals, and (3) HbA$_{1c}$ levels less than 7% for nonpregnant adults. Less stringent HbA$_{1c}$ goals may be appropriate in children, those with a history of severe hypoglycemia, limited life expectancy, and advanced microvascular and macrovascular disease. In the elderly frail patient, an HbA$_{1c}$ target of approximately 8% (preprandial blood glucose levels in the range of the 150-159 mg/dL) may be reasonable although formal evidence is lacking.

A. Type 1 diabetes   At the onset of type 1 diabetes, many patients recover some pancreatic β cell function and may temporarily need only low doses of exogenous insulin to supplement their own endogenous insulin secretion. This is known as the *honeymoon period*. Within 8 weeks to 2 years, however, most of these patients show either absent or negligible pancreatic β cell function. At this point, these patients should be switched to a more flexible insulin regimen with a combination of rapid-acting insulin analogs or regular insulin together with intermediate-acting or long-acting insulin. At a minimum, the patient should be on a three-injection regimen and frequently may need four or more injections. Twice-daily split-dose insulin mixtures cannot maintain near-normalization of blood glucose without hypoglycemia (particularly at night) and are not recommended. Self-monitoring of blood glucose levels is required for determining the optimal adjustment of insulin dosage and the modulation of food intake and exercise in type 1 diabetes.

TABLE 17–15 Examples of intensive insulin regimens using rapid-acting insulin analogs (insulin lispro, aspart, or glulisine) and NPH, or insulin detemir, glargine or degludec in a 70-kg man with type 1 diabetes.[a-c]

	Prebreakfast	Prelunch	Predinner	Bedtime
Rapid-acting insulin analog	5 U	4 U	6 U	—
NPH insulin	3 U	3 U	2 U	8-9 U
OR				
Rapid-acting insulin analog	5 U	4 U	6 U	—
Insulin glargine or degludec	—	—	—	15-16 U
Insulin detemir	6-7 U	—	—	8-9 U

[a]Assumes that patient is consuming approximately 75 g carbohydrate at breakfast, 60 g at lunch, and 90 g at dinner.

[b]The dose of rapid-acting insulin analogs can be raised by 1 or 2 U if extra carbohydrate (15-30 g) is ingested or if premeal blood glucose is >170 mg/dL. The rapid-acting insulin analogs can be mixed in the same syringe with NPH insulin.

[c]Insulin glargine or insulin detemir must be given as a separate injection.

A combination of rapid-acting insulin analogs and long-acting insulins (insulin glargine or insulin detemir or insulin degludec) allows for more physiologic insulin replacement. In clinical studies, combinations of rapid-acting insulin analogs (insulin lispro or insulin aspart) with meals together with intermediate-acting (NPH) or longer-acting insulin (insulin glargine) for basal coverage have now been shown to have improved HbA$_{1c}$ values with less hypoglycemia when compared with a regimen of regular insulin with meals and NPH at night. Table 17–15 illustrates some regimens that might be appropriate for a 70-kg person with type 1 diabetes eating meals of standard carbohydrate intake and moderate to low fat content.

Multiple injections of NPH insulin can be mixed in the same syringe as the insulin lispro, insulin aspart, and insulin glulisine. Insulin glargine or degludec are usually given once in the evening to provide 24-hour coverage. There are occasional patients in whom insulin glargine does not seem to last for 24 hours, and in such cases it needs to be given twice a day. Insulin detemir does not last for 24 hours and usually has to be given twice a day. The long acting insulin analogs cannot be *mixed* with any of the other insulins and must be given as a separate injection.

Continuous subcutaneous insulin infusion by portable battery-operated *open loop* insulin pumps currently provides the most flexible approach, allowing the setting of different basal rates throughout the 24 hours and permitting bolus dose adjustment by as little as 0.05 unit increments. The 24-hour basal dosage is

usually based on age and body weight. An adolescent might need as much as 0.4 U/kg/d; young adult (<25 years) 0.35 U/kg/d; and older adults 0.25 U/kg/d. For example, a 70-kg 30-year-old man may require a basal rate of 0.7 U/h throughout the 24 hours with the exception of 3 AM to 8 AM, when 0.8 U/h might be appropriate (to accomodate the *dawn phenomenon*). The dawn phenomenon is more pronounced in the adolescent and young adult. The meal bolus also varies based on the time of day and the person's age. Adolesecents and young adults usually require 1 unit for about 10 g carbohydrate. Older adults usually require about 1 unit for 15 g carbohydrate. The correction factor—how much insulin is needed to lower glucose levels by 50 mg/dL can be calculated from the insulin-to-carbohydrate ratios. For example, if 1 unit is required for 15 g carbohydrate, then 1 unit will lower glucose levels by 50 mg/dL. If on the other 1.5 units of insulin is required for 15 g of carbohydrate (ie, 1 unit for 10 g carbohydrate), then 1.5 units of insulin will lower glucose levels by 50 mg/dL (ie, 1 unit will lower glucose level by 33 mg/dL). For a 70-kg 30-year-old man, 1 unit per 15 g of carbohydrate plus 1 U for 50 mg/dL of blood glucose above a target value (eg, 120 mg/dL) is a common starting point. Further adjustments to basal and bolus dosages would depend on the results of blood glucose monitoring. Most patients use the rapid-acting insulin analogs in the pumps.

One of the more difficult therapeutic problems in managing patients with type 1 diabetes is determining the proper adjustment of insulin dose when the early morning blood glucose level is high before breakfast (Table 17–16). Prebreakfast hyperglycemia is sometimes due to the *Somogyi effect*, in which nocturnal hypoglycemia evokes a surge of counterregulatory hormones to produce high blood glucose levels by 7 AM. However, a more common cause of prebreakfast hyperglycemia is the waning of the evening or bedtime insulin and/or the dawn phenomenon. That is, reduced tissue sensitivity to insulin between 5 AM and 8 AM (dawn), due to spikes of growth hormone released hours before, at onset of sleep. Table 17–16 shows that diagnosis of the cause of prebreakfast hyperglycemia can be facilitated by self-monitoring of blood glucose at 3 AM in addition to the usual bedtime and 7 AM measurements. This is required for only a few nights, and when a particular pattern emerges from monitoring blood glucose levels overnight, appropriate therapeutic measures can be taken.

Prebreakfast hyperglycemia due to the Somogyi effect can be treated by reducing the dose of either intermediate- or long-acting insulin analog at bedtime. For hyperglycemia due to waning of overnight basal insulin and/or dawn phenomenon, an increase in the evening dose of the basal insulin or shifting it from dinnertime to bedtime (or both) can be effective. A bedtime dose either of insulin glargine or insulin detemir provides more sustained overnight insulin levels than human NPH and may be effective in managing refractory prebreakfast hyperglycemia. If this fails, insulin pump therapy may be required with a higher basal insulin infusion rate (eg, from 0.8 U/h to 0.9 U/h from 6 AM until breakfast).

Metformin use in type 1 diabetes patients has been shown to reduce the total daily insulin doses and promote weight loss but does not improve glucose control. GLP-1 receptor agonists similarly have been demonstrated to promote weight loss and reduce insulin doses in type 1 patients but not improve glucose control.

B. Type 2 diabetes   Therapeutic recommendations are based on the relative contributions of β cell insufficiency and insulin insensitivity in individual patients. The possibility that the individual patient has a specific etiological cause for their diabetes should always be considered, especially when the patient does not have a family history of type 2 diabetes or does not have any evidence of central obesity or insulin resistance. Such patients should be evaluated for other types of diabetes such as LADA or MODY. Patients with LADA should be prescribed insulin when their disease is diagnosed and treated like patients with type 1 diabetes. It is also important to note that many type 2 patients have a progressive loss of β cell function and will require additional therapeutic interventions with time.

Weight reduction—One of the primary modes of therapy in the obese patient with type 2 diabetes is weight reduction. Normalization of glycemia can be achieved by reducing adipose stores, with consequent restoration of tissue sensitivity to insulin. A combination of caloric restriction, increased exercise, modification of behavior, and consistent reinforcement of good eating habits is required if a weight reduction program is to be successful. Understanding the risks and complications of diabetes may motivate the patient to lose weight.

TABLE 17–16 Typical patterns of overnight blood glucose levels and serum-free immunoreactive insulin levels in prebreakfast hyperglycemia due to various causes in patients with type 1 diabetes.

	Blood Glucose Levels (mg/dL)			Serum-Free Immunoreactive Insulin Levels (µU/mL)		
	10 PM	3 AM	7 AM	10 PM	3 AM	7 AM
Somogyi effect	90	40	200	High	Slightly high	Normal
Dawn phenomenon	110	110	150	Normal	Normal	Normal
Waning of circulating insulin levels plus dawn phenomenon	110	190	220	Normal	Low	Low
Waning of circulating insulin levels plus dawn phenomenon plus Somogyi effect	110	40	380	High	Normal	Low

For selected patients, medical or surgical options for weight loss should be considered (also see Chapter 20). Orlistat, phentermine/topiramate, naltrexone/extended-release bupropion, and high-dose liraglutide (3 mg daily) are weight loss medications approved for use in combination with diet and exercise.

Orlistat is a reversible inhibitor of gastric and pancreatic lipases and prevents the hydrolysis of dietary triglycerides. These triglycerides are then excreted in the feces. In a 1-year study in obese patients with type 2 diabetes, those taking orlistat had lost more weight, had lower HbA$_{1c}$ values, and had improved lipid profiles. The main adverse reactions were gastrointestinal, with oily spotting, oily stool, flatus, and fecal urgency and frequency. Malabsorption of fat-soluble vitamins also occurs. Patients should take a multivitamin tablet containing fat-soluble vitamins at least 2 hours before or 2 hours after the administration of orlistat. Cases of severe liver injury have been reported with this drug although a cause and effect relationship has not been established.

Phentermine is a sympathomimetic amine that stimulates release of norepinephrine in the hypothalamus. Topiramate is primarily used as an anticonvulsant, but it also appears to reduce appetite. In a 56-week phase 3 study, an extended-release preparation of phentermine/topiramate (Qsymia), together with diet and lifestyle intervention, resulted in 10 kg weight loss (9.8%) compared to 1.4 kg (1.2%) with placebo. As might be expected, the diabetes subgroup on active therapy had greater reductions in HbA$_{1c}$ levels and fewer patients with prediabetes on active therapy progressed to diabetes. The adverse events are consistent with those of the constituent drugs. The most common adverse reactions were paresthesia, dizziness, dysgeusia, insomnia, constipation, and dry mouth. Topiramate can worsen depression and increase risk of suicidal thoughts. It is also teratogenic and the FDA has required the manufacturer to conduct a risk evaluation and mitigation strategy (REMS). The medication is only available through specialty mail-order pharmacies.

Lorcaserin (Belviq) is a 5-hydroxytryptamine receptor subtype 2C (5-HT2C) agonist. This receptor subclass regulates mood and appetite. In a 52-week study, patients taking lorcaserin had a 8.1-kg weight loss (8.2%) compared to 3.2 kg in the placebo group (3.3%). The main adverse reactions were headache and nausea. Fenfluramine, an agonist for the 5-HT2B receptor, was associated with serotonin-related cardiac valvulopathy. Activation of the 5-HT2C receptor, however, does not appear to be associated with valvulopathy. Sibutramine, a combined serotonin-norepinephrine reuptake inhibitor, was moderately effective in promoting weight loss, but it was withdrawn from the U.S. market because of its association with increased cardiovascular risk.

Naltrexone is an antagonist at mu- and kappa-opioid receptors and is used in the treatment of alcohol and opioid dependence. Bupropion is a partial agonist at the mu-opioid receptor, an antagonist at the kappa-opioid receptor and a partial agonist at the nociception receptor; it is used to treat depression, seasonal affective disorder, and as an aid to stop smoking. Naltrexone/extended-release bupropion (Contrave), together with diet and exercise in patients with diabetes, resulted in 2% more weight loss than placebo at 1 year of study. The weight loss with the medication was greater in the study of obese people without diabetes (4.1%). Serious neuropsychiatric events and seizures have been reported in patients taking bupropion. Naltrexone should not be given to patients receiving long-term opioid therapy. Naltrexone/extended bupropion should be discontinued if the patient needs intermittent opiate therapy. Patients should be warned that they may be more sensitive to opiates after the naltrexone/extended bupropion preparation has been discontinued.

Liraglutide 3 mg (Saxenda) is a GLP-1 receptor agonist. The 0.6 to 1.8 mg dose has been approved for the treatment of type 2 diabetes and is associated with modest weight loss. The 3-mg dose has been approved for weight loss in combination with diet and exercise. In nondiabetic obese persons, liraglutide together with diet and exercise resulted in 4.5% more weight loss than placebo at 1 year of treatment. In a study of people with diabetes, the average weight loss with the medication was 3.7% compared to placebo at 1 year. Common adverse reactions include nausea, vomiting, and diarrhea. Serious side effects include pancreatitis. The medication should not be used in patients with MEN 2 or personal or family history of medullary thyroid cancer.

Bariatric surgery (Roux-en-Y, gastric sleeve, biliopancreatic diversion/duodenal switch, or gastric banding) typically result in substantial weight loss, and improvement in glucose levels. A meta-analysis examining the impact of bariatric surgery on patients with diabetes and BMI of 40 kg/m^2 or greater noted that 82% of patients had resolution of clinical and laboratory manifestations of diabetes in the first 2 years after surgery and 62% remained free of diabetes more than 2 years after surgery. The improvement was most marked in the procedure that caused the greatest weight loss (biliopancreatic diversion/duodenal switch). There was, however, a high attrition of patients available for follow-up, and there was little information about different ethnic types. Weight regain does occur after bariatric surgery, and it can be expected that 20% to 25% of the lost weight will be regained over 10 years. The impact of this weight gain on diabetes recurrence depends principally on the degree of beta cell dysfunction. Also anatomic changes imposed by malabsorptive surgery can result in protein malnutrition, vitamin, and mineral deficiencies. Clinically significant deficiencies in calcium, folic acid, iron, and vitamins D, B$_{12}$, A, and K are common. Thus, patients undergoing malabsorptive procedures require lifelong supplementation and monitoring by a team familiar with possible deficiencies. Both early and late dumping symptoms can also occur.

Nonobese patients with type 2 diabetes frequently have increased visceral adiposity—the so-called metabolically obese normal-weight patient. There is less emphasis on weight loss but exercise remains an important aspect of treatment.

Antihyperglycemic agents—Figure 17–13 outlines the treatment approach based on the consensus algorithm proposed by the American Diabetes Association and the European Association for the Study of Diabetes. The current recommendation is to start metformin therapy at diagnosis and not wait to see if the patient can achieve target glycemic control with weight management and exercise. **Metformin** is advantageous because, apart from lowering glucose without the risk of hypoglycemia, it also lowers triglycerides and promotes some modest weight loss. The drug, however,
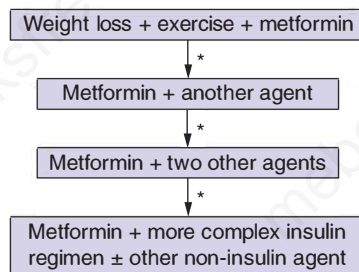
*If needed to reach individualized HbA_{1c} target after ~3 months

Seven main classes of agents: metformin, sulfonylureas (includes nateglinide, repaglinide), thiazolidinediones, GLP-1 receptor agonists, DPP-4 inhibitors, SGLT2 inhibitors, insulins (alpha-glucosidase inhibitors, colesevelam, pramlintide, bromocriptine not included because of limited efficacy and significant adverse reactions).

In making a therapeutic decision (single agent or in combination) take into account efficacy; hypoglycemia risk; effect on weight; major side effects; cost. The DPP-4 inhibitors are of moderate efficacy whereas all the other agents are of high efficacy. Sulfonylureas and insulins have increased risk of hypoglycemia. Metformin and DPP-4 inhibitors are weight neutral; GLP-1 receptor agonists, SGLT2 inhibitors promote weight loss; and sulfonylureas, insulins and thiazolidinediones are associated with weight gain. Metformin can cause lactic acidosis. Thiazolidinediones are associated with fluid retention, fracture risk and possibly bladder cancer. GLP-1 receptor agonists are associated with nausea and vomiting and possibly pancreatitis. DPP-4 inhibitors may increase pancreatitis risk. SGLT inhibitors are associated with genital mycotic infections, urinary tract infections, and hypotension. All the agents with the exception of metformin and sulfonylureas are expensive. Insulins are expensive if you take into consideration the additional costs of monitoring.

FIGURE 17–13 Suggested algorithm for the treatment of type 2 diabetes based partly on the recommendations of the consensus panel of the American Diabetes Association/European Association for the Study of Diabetes. (Data from Inzucchi SE, Bergenstal RM, Buse JB, et al. Management of hyperglycaemia in type 2 diabetes, 2015: a patient-centred approach. Update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetologia*. 2015 Mar;58(3):429-442.)

cannot be used in patients with renal or liver failure, and some patients have gastrointestinal side effects at even the lowest doses. Under these circumstances the choice of the initial agent depends on a number of factors, including comorbid conditions, adverse reactions to the medications, ability of the patient to monitor for hypoglycemia, drug cost, and patient and physician preferences.

Sulfonylureas have been available for many years, and their use in combination with metformin is well established. They do, however, have the propensity to cause hypoglycemia and weight gain. **Thiazolidinediones** improve peripheral insulin resistance and lower glucose without causing hypoglycemia. They also have been reported to improve nonalcoholic fatty liver disease. In addition, they have beneficial effects on the lipid profile and some other cardiovascular risk factors. They decrease microalbuminuria, and reduce neointimal tissue hyperplasia after coronary artery stent placement. Thiazolidinediones, however, can cause fluid retention and are contraindicated in patients with heart failure. They also very commonly increase weight, which patients find distressing, affecting adherence. The drugs are associated with increased fracture risk in women and this adverse effect significantly limits their use. Both drugs are contraindicated in patients with active liver disease and in patients with liver enzyme levels more than or equal to 2.5 times the upper limit of normal. The

α-glucosidase inhibitors have modest glucose-lowering effects and have gastrointestinal side effects. They have a lower risk of hypoglycemia than the sulfonylureas and promote weight loss. The GLP-1 receptor agonists have a lower risk of hypoglycemia than the sulfonylureas, and they promote weight loss. However, they need to be given by injection, cause nausea, may cause pancreatitis, and are contraindicated in patients with gastroparesis. The DPP-4 inhibitors also have a low risk of hypoglycemia, and they do not cause nausea or vomiting. They can also be used in patients with kidney impairment. There are, however, reports of serious allergic reactions, including anaphylaxis, angioedema, and Stevens-Johnson syndrome. There is concern that they, like the GLP-1 receptor agonists, may cause pancreatitis. The SGLT2 inhibitors lower fasting and postprandial glucose levels. They also have a low risk of hypoglycemia, promote weight loss, and lower blood pressure levels. They increase the risk for mycotic genital infections, urosepsis and possibly fractures. They can also cause volume depletion and are less effective in patients with kidney disease.

When patients are not well controlled on their initial therapy (usually metformin) then a second agent should be added. In those patients where the problem is hyperglycemia after a carbohydrate-rich meal (such as dinner), a short-acting secretagogue

before meals may suffice to put the glucose levels into the target range. Patients with severe insulin resistance or nonalcoholic fatty liver disease or microalbuminuria may be candidates for a thiazolidinedione. Subjects who are very concerned about weight gain may benefit from a trial of a GLP-1-receptor agonist, DPP-4 inhibitor, or SGLT2 inhibitor. If two agents are inadequate, then a third agent is added, although data on efficacy with such combined therapy are limited. When the combination of oral agents (and injectable GLP-1 receptor agonists) fail to achieve target glycemic control in patients with type 2 diabetes or if there are contraindications to their use, then insulin treatment should be instituted. Various insulin regimens may be effective. One proposed regimen is to continue the oral combination therapy and then simply add a bedtime dose of NPH or long-acting insulin analog to reduce excessive nocturnal hepatic glucose output and improve fasting glucose levels. If the patient does not achieve target glucose levels during the day, then daytime insulin treatment can be initiated. A convenient insulin regimen under these circumstances is a split dose of 70/30 NPH/regular mixture (or Humalog Mix 75/25 or NovoLogMix 70/30) before breakfast and before dinner. If this regimen fails to achieve satisfactory glycemic goals or is associated with unacceptable frequency of hypoglycemic episodes, then a more intensive regimen of multiple insulin injections can be instituted as in patients with type 1 diabetes. Metformin principally reduces hepatic glucose output, and it is reasonable to continue with this drug when insulin therapy is instituted. The thiazolidinediones, which improve peripheral insulin sensitivity, can be used together with insulin, but this combination is associated with more weight gain and peripheral edema. The sulfonylureas, the GLP-1-receptor agonists, and the DPP-4 inibitors also have been shown to be of continued benefit. Weight-reducing interventions should continue even after initiation of insulin therapy and may allow for simplification of the therapeutic regimen in the future.

IMMUNOPATHOLOGY OF INSULIN THERAPY

At least five molecular classes of insulin antibodies are produced during the course of insulin therapy: IgA, IgD, IgE, IgG, and IgM. Human insulin is much less antigenic than the older formulations of animal (especially beef) insulins, but because of its hexameric presentation at therapeutic injection doses, it is also treated as a foreign substance by the immune system and results in detectable—albeit low—titers of insulin antibodies in most patients.

A. Insulin allergy   Insulin allergy, a hypersensitivity reaction of the immediate type, is a rare condition in which local or systemic urticaria occurs immediately after insulin injection. This reaction is due to histamine release from tissue mast cells sensitized by adherence of IgE antibodies to their surface. In severe cases, anaphylaxis can occur. The appearance of a subcutaneous nodule at the site of insulin injection, occurring several hours after the injection and lasting for up to 24 hours, has been attributed to an IgG-mediated complement-binding Arthus reaction. The

sensitivity can be to the insulin or to the additives–protamine or cresol. Because sensitivity was often due to noninsulin protein contaminants, the highly purified insulins have markedly reduced the incidence of insulin allergy, especially of the local variety. Antihistamines, corticosteroids, and even desensitization may be required, especially for systemic hypersensitivity in an insulin-dependent patient. Rituximab followed by Omalizumab was reported to be successful in one patient with severe refractory insulin allergy. A trial of insulin analogs should also be considered. There is a case report of successful use of insulin lispro in the face of generalized allergy to human insulin.

B. Immune insulin resistance   All patients who receive insulin (including insulin analogs) develop a low titer of circulating IgG antibodies, and this neutralizes the rapid action of insulin to a small extent. With the animal insulins, a high titer of circulating antibodies is sometimes developed, resulting in extremely high insulin requirements, often to more than 200 U/d. This is now very rarely seen with the switch to the highly purified human insulins and has not been reported with use of the analogs.

C. Lipodystrophy at injection sites   Rarely, a disfiguring atrophy of subcutaneous fatty tissue occurs at the site of insulin injection. Although the cause of this complication is obscure, it seems to represent a form of immune reaction, particularly because it occurs predominantly in females and is associated with lymphocyte infiltration in the lipoatrophic area. This complication has become even less common because of the development of highly purified insulin preparations of neutral pH. Injection of highly purified preparations of insulin directly into the atrophic area often results in restoration of normal contours.

Lipohypertrophy, on the other hand, is not a consequence of immune responses; rather, it seems to be due to the pharmacologic effects of depositing insulin in the same location repeatedly (Figure 17–14). It is prevented by rotation of injection sites. There is a

FIGURE 17–14 Lipohypertrophy due to overuse of insulin infusion sites.

case report of a male patient who had intractable lipohypertrophy (fatty infiltration of injection site) with human insulin but no longer had the problem when he switched to insulin lispro.

ACUTE COMPLICATIONS OF DIABETES MELLITUS

Hypoglycemia

Hypoglycemic reactions (see Chapter 18) are the most common complications that occur in patients with diabetes who are treated with insulin. It can also occur in any patient taking oral agents that stimulate pancreatic β cells (eg, sulfonylureas, meglitinide, D-phenylalanine analogs), particularly if the patient is elderly and has renal or liver disease. It occurs more frequently with the use of long-acting sulfonylureas.

The signs and symptoms of hypoglycemia may be divided into those resulting from stimulation of the autonomic nervous system and those arising from neuroglycopenia (insufficient glucose for normal central nervous system function). When the blood glucose falls to around 54 mg/dL (3 mmol/L), the patient starts to experience both sympathetic (tachycardia, palpitations, sweating, tremulousness) and parasympathetic (nausea, hunger) nervous system symptoms. If these autonomic symptoms are ignored and the glucose levels fall further (to around 50 mg/dL [2.8 mmol/L]), then neuroglycopenic symptoms appear, including irritability, confusion, blurred vision, tiredness, headache, and difficulty speaking. A further decline in glucose (below 30 mg/dL [1.7 mmol/L]) can then lead to loss of consciousness or even a seizure.

With repeated episodes of hypoglycemia, there is adaptation and autonomic symptoms do not occur until the blood glucose levels are much lower and so the first symptoms are often due to neuroglycopenia. This condition, which is referred to as **hypoglycemic unawareness**, results from failure of the sympathetic nervous system to respond to hypoglycemia (Table 17–17). This adaptation of the central nervous system to recurrent hypoglycemic episodes is due to changes in glucose transportation or metabolism. It has been shown that hypoglycemic unawareness can be reversed by keeping glucose levels high for a period of several weeks. Except for sweating, most of the sympathetic symptoms of hypoglycemia are blunted in patients receiving β-blocking agents for angina pectoris or hypertension. Though not absolutely contraindicated, these drugs must be used with caution in insulin-requiring diabetics, and β₁-selective blocking agents are preferred.

Hypoglycemia in insulin-treated patients with diabetes occurs as a consequence of three factors: behavioral issues, impaired counterregulatory systems, and complications of diabetes.

Behavioral issues include injecting too much insulin for the amount of carbohydrates ingested. Drinking alcohol in excess, especially on an empty stomach, can also cause hypoglycemia. In patients with type 1 diabetes, hypoglycemia can occur during or even several hours after exercise, and so glucose levels need to be monitored and food and insulin adjusted. Some patients do not like their glucose levels to be high, and they treat every high glucose level aggressively. These individuals who *stack* their insulin, that is, give another dose of insulin before the first injection has had its full action, can develop hypoglycemia.

Counterregulatory issues resulting in hypoglycemia include impaired glucagon response and impaired sympathoadrenal responses (Table 17–18). Patients with diabetes of greater than 5 years duration lose their glucagon response to hypoglycemia. As a result, they are at a significant disadvantage in protecting themselves against falling glucose levels. Once the glucagon response is lost, their sympathoadrenal responses take on added importance. Unfortunately, aging, autonomic neuropathy, or hypoglycemic unawareness due to repeated low glucose levels further blunts the sympathoadrenal responses. Occasionally, Addison disease develops in persons with type 1 diabetes mellitus; when this happens, insulin requirements fall significantly, and unless insulin dose is reduced, recurrent hypoglycemia will develop.

TABLE 17–17 Hypoglycemic "unawareness" in type 1 diabetes mellitus.

I. Sleeping patient (nocturnal hypoglycemia)
II. Hypoglycemia with unawareness while awake
   A. Manifestations
     a. Neuroglycopenic symptoms first (weakness, lethargy, confusion, incoordination, blurred vision)
     b. Autonomic symptoms are delayed and blunted (tremor, anxiety, palpitations, sweating, hunger)
   B. Mechanisms
     1. Defective autonomic response:
       a. Adaptation to chronic hypoglycemia (increased brain glucose transporter I)
       b. Due to diabetic autonomic neuropathy
   C. Management
     1. Identify patients at risk and reevaluate glycemic goals
     2. Advise frequent self-monitoring of blood glucose; use of continuous glucose monitoring systems
     3. Learn to detect subtle symptoms of neuroglycopenia
     4. Avoid recurrent hypoglycemia
     5. Injectable glucagon made available to family

TABLE 17–18 Counterregulatory responses to hypoglycemia.

Normal Counterregulation	Defective Counterregulation in Type 1 Diabetes[a]
Glucagon rises rapidly to three to five times baseline after insulin-induced hypoglycemia, provoking hepatic glycogenolysis	Glucagon response to insulin-induced hypoglycemia is lost after onset of type 1 diabetes
Adrenergic discharge (1) raises hepatic glucose output by glycogenolysis and (2) provides warning to subject of impending hypoglycemic crisis	Blunted or absent adrenergic response may occur as a result of: (1) Neural damage associated with advanced age or autonomic neuropathy (2) Neural dysfunction (iatrogenic) from frequent hypoglycemia

[a]Type 2 diabetics are less well characterized as to their defective counterregulation of glucagon loss, but appear to have the same frequency and causes of adrenergic loss as do type 1 diabetics.

Complications of diabetes that increase the risk for hypoglycemia include autonomic neuropathy, gastroparesis, and renal failure. The sympathetic nervous system is an important system alerting the individual that the glucose level is falling by causing symptoms of tachycardia, palpitations, sweating, and tremulousness. Failure of the sympathoadrenal responses increases the risk of hypoglycemia. In patients with gastroparesis, insulin given before a meal promotes maximal glucose uptake into cells before the food is absorbed, causing the glucose levels to fall. Finally, in renal failure, hypoglycemia can occur presumably because of decreased insulin clearance as well as loss of the renal contribution to gluconeogenesis in the postabsorptive state.

To treat insulin-induced hypoglycemia, the diabetic patient should carry glucose tablets or juice at all times. For most episodes, ingestion of 15 g of carbohydrate is sufficient to reverse the hypoglycemia. The patient should be instructed to check the blood glucose in 15 minutes and treat again if the glucose level is still low.

A parenteral glucagon emergency kit (1 mg) should be provided to every patient with diabetes who is on insulin therapy. Family or friends should be instructed how to inject it subcutaneously or intramuscularly into the buttock, arm, or thigh in the event that the patient is unconscious or refuses food. The drug can occasionally cause vomiting and the unconscious patient should be turned on his or her side to protect the airway. Glucagon mobilizes glycogen from the liver raising the blood glucose by about 36 mg/dL (2 mmol/L) in about 15 minutes. After the patient recovers consciousness additional oral carbohydrate should be given. Glucagon is contraindicated in sulfonylurea-induced hypoglycemia where it paradoxically causes insulin release. People with diabetes on hypoglycemic drug therapy should also wear a MedicAlert bracelet or necklace or carry a card in his or her wallet.

Medical personnel treating severe hypoglycemia can give 50 mL of 50% glucose solution by rapid intravenous infusion. If intravenous therapy is not available, 1 mg of glucagon can be injected intramuscularly. If the patient is stuporous and glucagon is not available, small amounts of honey or maple syrup or glucose gel (15 g) can be inserted within the buccal pouch, although, in general, oral feeding is contraindicated in unconscious patients. Rectal administration of maple syrup or honey (30 mL per 500 mL of warm water) has been effective.

Most patients who arrive at emergency departments in hypoglycemic coma appear to recover fully; however, profound hypoglycemia or delays in therapy can result in permanent neurologic deficit or even death. Furthermore, repeated episodes of hypoglycemia may have a cumulative adverse effect on intellectual functioning. The physician should carefully review with the patient the events leading up to the hypoglycemic episode. Associated use of other medications, as well as alcohol or narcotics, should be noted. Careful attention should be paid to diet, exercise pattern, insulin or sulfonylurea dosage, and general compliance with the prescribed diabetes treatment regimen. Any factors thought to have contributed to the development of the episode should be identified and recommendations made in order to prevent recurrences of this potentially disastrous complication of diabetes therapy.

If the patient is hypoglycemic from use of a long-acting oral hypoglycemic agent (eg, chlorpropamide or glyburide) or from high doses of a long-acting insulin, admission to hospital for treatment with continuous intravenous glucose and careful monitoring of blood glucose is indicated.

Coma

Coma is a *medical emergency* calling for immediate evaluation to determine its cause so that proper therapy can be started. Patients with diabetes may be comatose because of hypoglycemia, diabetic ketoacidosis, hyperglycemic hyperosmolar coma, or lactic acidosis. When evaluating a comatose diabetic patient, these must be considered *in addition* to the myriad causes included in the differential diagnosis of coma (eg, cerebrovascular accidents, head trauma, intoxication with alcohol, or other drugs).

After emergency measures have been instituted (airway protection; laboratory tests; intravenous dextrose unless fingerstix blood glucose shows hyperglycemia), a careful history (from family, friends, or paramedics), physical examination, and laboratory evaluation are required to resolve the differential diagnosis. Patients in deep coma from a hyperosmolar nonketotic state or from hypoglycemia are generally flaccid and have quiet breathing—in contrast to patients with acidosis, whose respirations are rapid and deep if the pH of arterial blood has dropped to 7.1 or below. When hypoglycemia is a cause of the coma, the state of hydration is usually normal. Although the clinical laboratory remains the final arbiter in confirming the diagnosis, a rapid *estimate* of blood glucose and ketones can be obtained by the use of bedside glucose and ketone meters (see Laboratory Findings in diabetes mellitus, discussed earlier). Table 17–19 is a summary of some laboratory abnormalities found in diabetic patients with coma attributable to diabetes or its treatment.

1. DIABETIC KETOACIDOSIS

This acute complication of diabetes mellitus may be the first manifestation of previously undiagnosed type 1 diabetes or may result from increased insulin requirements in type 1 diabetes patients during the course of infection, trauma, myocardial infarction, or surgery. The National Data Group reports an annual incidence of five to eight episodes of diabetic ketoacidosis per 1000 diabetic patients. In all cases, precipitating factors such as infection should be searched for and treated appropriately. Poor compliance, either for psychological reasons or because of inadequate patient education, is probably the most common cause of recurrent diabetic ketoacidosis.

Diabetic ketoacidosis has been found to be one of the more common serious complications of insulin pump therapy, occurring in approximately 1 per 80 patient-months of treatment.

Patients with type 2 diabetes may also develop ketoacidosis under severe stress such as sepsis, trauma, or major surgery.

Pathogenesis

Acute insulin deficiency results in rapid mobilization of energy from stores in muscle and fat depots, leading to an increased flux

TABLE 17–19 Summary of some laboratory abnormalities in patients with coma directly attributable to diabetes or its treatment.

	Urine		Plasma			
	Glucose	**Ketones**	**Glucose**	**Bicarbonate**	**Ketones**	**Osmolality**
Diabetic ketoacidosis	++ to ++++	++++	High	Low	++++	+++
Hyperglycemic nonketotic coma	++ to ++++	0 or +[a]	High	Normal or slightly low[b]	0	++++
Hypoglycemia	0[c]	0 or +	Low	Normal	0	Normal
Lactic acidosis	0 to +	0 or +	Normal, low, or high	Low	0 or +	Normal

[a]A small degree of ketonuria may be present if the patient is severely stressed or has not been eating because of illness.

[b]A patient may be acidotic if there is severe volume depletion with cardiovascular collapse or if sepsis is present.

[c]Leftover urine in bladder might still contain sugar from earlier hyperglycemia.

of amino acids to the liver for conversion to glucose and of fatty acids for conversion to ketones (acetoacetate, β-hydroxybutyrate, and acetone). In addition to this increased availability of precursor, there is a direct effect of the low insulin-glucagon ratio on the liver that promotes increased production of ketones as well as of glucose. In response to both the acute insulin deficiency and the metabolic stress of ketosis, the levels of insulin-antagonistic hormones (corticosteroids, catecholamines, glucagon, and GH) are consistently elevated. Furthermore, in the absence of insulin, peripheral utilization of glucose and ketones is reduced. The combination of increased production and decreased utilization leads to an accumulation of these substances in blood, with plasma glucose levels reaching 500 mg/dL (27.8 mmol/L) or more and plasma ketones reaching levels of 8 to 15 mmol/L or more. β-Hydroxybutyrate is the predominant ketone and its ratio to acetoacetate increases from 1:1 to as much as 5:1.

The hyperglycemia causes osmotic diuresis leading to depletion of intravascular volume. As this progresses, impaired renal blood flow reduces the kidney's ability to excrete glucose, and hyperosmolality worsens. Severe hyperosmolality (>330 mOsm/kg) correlates closely with central nervous system depression and coma.

In a similar manner, impaired renal excretion of hydrogen ions aggravates the metabolic acidosis that occurs as a result of the accumulation of the ketoacids, β-hydroxybutyrate, and acetoacetate. The accumulation of ketones may cause vomiting, which exacerbates the intravascular volume depletion. In addition, prolonged acidosis can compromise cardiac output and reduce vascular tone. The result may be severe cardiovascular collapse with generation of lactic acid, which then adds to the already existent metabolic acidosis.

Clinical Features

A. Symptoms and signs

The appearance of diabetic ketoacidosis is usually preceded by a day or more of polyuria and polydipsia associated with marked fatigue, nausea, and vomiting. Eventually, mental stupor ensues and can progress to frank coma. On physical examination, evidence of dehydration in a stuporous

patient with rapid and deep respirations and the *fruity* breath odor of acetone strongly suggest the diagnosis. Postural hypotension with tachycardia indicates profound dehydration and salt depletion. Abdominal pain and even tenderness may be present in the absence of abdominal disease, and mild hypothermia is usually present.

B. Laboratory findings

Typically, the patient with moderately severe diabetic ketoacidosis has a plasma glucose of 350 to 900 mg/dL (19.4-50 mmol/L), serum ketones are positive at a dilution of 1:8 or greater and β-hydroxybutyrate levels are 4 mmol/L or higher, hyperkalemia of 5 to 8 mEq/L, slight hyponatremia of approximately 130 mEq/L, hyperphosphatemia of 6 to 7 mg/dL, and an elevated blood urea nitrogen and creatinine. Acidosis may be severe (pH ranging from 6.9-7.2 with a bicarbonate concentration ranging from 5 to 15 mEq/L); pCO_2 is low (15-20 mm Hg) secondary to hyperventilation. The difference between venous and arterial pH is 0.02 to 0.15 pH units and the difference in venous and arterial bicarbonate is 1.88 mEq/L. These small differences will not affect either the diagnosis or the management of diabetic ketoacidosis, and there is no need to collect arterial blood for measuring the acid-base status. The fluid depletion is typically about 100 mL/kg.

The hyperkalemia occurs despite total body potassium depletion, because of the shift of potassium from the intracellular to extracellular spaces in systemic acidosis. The average total body potassium deficit resulting from osmotic diuresis, acidosis, and gastrointestinal losses is about 3 to 5 mEq/kg body weight. Similarly, despite the elevated serum phosphate, total body phosphate is generally depleted. Serum sodium is generally reduced, due to loss of sodium ions by polyuria and vomiting (7-10 mEq/kg), and because severe hyperglycemia shifts intracellular water into the extracellular compartment (for every 100 mg/dL of plasma glucose above normal, serum sodium decreases by 1.6 mEq/L [5.56 mmol/L]). Hypertriglyceridemia should be considered if the corrected sodium is very low. Serum osmolality can be directly measured by standard tests of freezing point depression or can be

estimated by calculating the molarity of sodium, chloride, and glucose in the serum. A convenient formula for estimating *effective* serum osmolality is:

$$mOsm/kg = 2[measured\ Na^+] + \frac{Glucose\ (mg/dL)}{18}$$

The effective serum osmolality in humans is generally between 280 and 300 mOsm/kg. These calculated estimates are usually 10 to 20 mOsm/kg lower than values recorded by standard cryoscopic techniques. Central nervous depression or coma occurs when the effective serum osmolality exceeds 320 to 330 mOsm/L.

Blood urea nitrogen and serum creatinine are invariably elevated because of dehydration. Urea exerts an effect on freezing point depression as measured in the laboratory, but it is freely permeable across cell membranes and therefore not included in calculations of effective serum osmolality. Serum creatinine may also be falsely elevated due to interference from acetoacetate with some automated creatinine assays. However, most laboratories can correct for these interfering chromogens by using a more specific method, if asked to do so.

The nitroprusside reagents (Acetest and Ketostix) used for the bedside assessment of ketoacidemia and ketoaciduria measure only acetoacetate and its by-product, acetone. The sensitivity of these reagents for acetone, however, is quite poor, requiring over 10 mmol/L. This level is seldom reached in the plasma of ketoacidotic subjects—although this detectable concentration is readily achieved in urine. Thus, in the plasma of ketotic patients, only acetoacetate is measured by these reagents. The more prevalent β-hydroxybutyrate has no ketone group and is therefore not detected by the conventional nitroprusside tests. This takes on special importance in the presence of circulatory collapse during diabetic ketoacidosis, wherein an increase in lactic acid can shift the redox state to increase β-hydroxybutyrate at the expense of the readily detectable acetoacetate. Bedside diagnostic reagents would then be unreliable, suggesting no ketonemia in cases where β-hydroxybutyrate is a major factor in producing the acidosis. Combined bedside glucose and ketone meters (Precision Xtra, Nova Max Plus) that measure blood β-hydroxybutyrate concentration on capillary blood are now available. Many clinical laboratories also offer direct blood β-hydroxybutyrate measurement.

Nonspecific elevations of serum amylase and lipase occur in about 16% to 25% of cases of diabetic ketoacidosis, and an imaging study may be necessary if the diagnosis of acute pancreatitis is being seriously considered. Leukocytosis as high as 25,000/mL with a left shift may occur with or without associated infection. The presence of an elevated or even a normal temperature would suggest the presence of an infection, since patients with diabetic ketoacidosis are generally hypothermic if uninfected.

Treatment

Patients with mild DKA are alert and have pH between 7.25 and 7.30 and β-hydroxybutyrate levels of 3 to 4 mmol/L; those with moderate DKA are alert or slightly drowsy and have pH between 7.0 and 7.24 and β-hydroxybutyrate levels of 4 to 8 mmol/L; and those with severe DKA are stuporous, have pH less than 7.0 and β-hydroxybutyrate levels of greater than 8 mmol/L. Those with mild DKA can be treated in the emergency room, but those with moderate or severe DKA require admission to the intensive care unit or step-down unit.

The therapeutic goals are to restore plasma volume and tissue perfusion; reduce blood glucose and osmolality toward normal; correct acidosis; replenish electrolyte losses; and identify and treat precipitating factors. Gastric intubation is recommended in the comatose patient to prevent vomiting and aspiration that may occur as a result of gastric atony, a common complication of diabetic ketoacidosis. In patients with preexisting cardiac or renal failure or those in severe cardiovascular collapse, a central venous pressure catheter or a Swan-Ganz catheter should be inserted to evaluate the degree of hypovolemia and to monitor subsequent fluid administration.

A comprehensive flow sheet that includes vital signs, serial laboratory data, and therapeutic interventions (eg, fluids, insulin) should be meticulously maintained by the clinician responsible for the patient's care. Plasma glucose should be recorded hourly and electrolytes and pH at least every 2 to 3 hours during the initial treatment period. A bedside glucose meter should be used to titrate the insulin therapy. The patient should not receive sedatives or opioids in order to avoid masking signs and symptoms of impending cerebral edema.

1. **Fluid replacement.** In most adult patients, the fluid deficit is 4 to 5 L. Once the diagnosis of diabetic ketoacidosis is established in the emergency department, administration of at least 2 L of isotonic saline (0.9% saline solution) in an adult patient in the first 2 to 3 hours is necessary to help restore plasma volume and stabilize blood pressure while acutely reducing the hyperosmolar state. In addition, by improving renal plasma flow, fluid replacement also restores the renal capacity to excrete hydrogen ions, thereby ameliorating the acidosis as well. After the first 2 L of fluid have been given, the intravenous infusion should be at a rate of 300 to 400 mL/h. Use 0.9% saline unless the serum sodium is greater than 150 mEq/L when 0.45% ("half normal") saline should be used. Failure to give sufficient volume replacement (at least 3-4 L in 8 hours) to restore normal perfusion is one of the most serious therapeutic shortcomings affecting satisfactory recovery. In the same way, excessive fluid replacement (>5 L in 8 hours) may contribute to acute respiratory distress syndrome or cerebral edema. When blood glucose falls to approximately 250 mg/dL (13.9 mmol/L), the fluids should be changed to a 5% glucose containing solution to maintain plasma glucose in the range of 250 to 300 mg/dL (13.9-16.7 mmol/L). This prevents the development of hypoglycemia and also reduces the likelihood of cerebral edema, which may result from a too rapid decline of blood glucose.

2. **Insulin treatment.** Immediately after the initiation of fluid replacement and determination that the patient's serum potassium is more than 3.5 mEq/L, a bolus of 0.1 U/kg can be given intravenously to prime the tissue insulin receptors. This inhibits both gluconeogenesis and ketogenesis while promoting utilization of glucose and keto acids. Following the initial bolus, an intravenous insulin infusion is initiated at a rate of 0.1 U/kg/h. A prospective randomized study showed that a bolus dose is not required if patients are given insulin infusion

at a rate of 0.14 U/kg/h. When a continuous infusion of insulin is used, 25 U of regular human insulin should be placed in 250 mL of isotonic saline and the first 50 mL of solution flushed through to saturate the tubing before connecting it to the intravenous line. The insulin infusion should be piggy-backed into the fluid line so that the rate of fluid replacement can be changed without altering the insulin delivery rate. If the plasma glucose level fails to fall at least 10% in the first hour, a repeat loading dose (0.1 or 0.14 U/kg) is recommended. Rarely, a patient with insulin resistance is encountered; this requires doubling the insulin dose every 2 to 4 hours if severe hyperglycemia does not improve after the first two doses of insulin and fluid replacement. The insulin dose should be adjusted with the goal of lowering the glucose concentration by about 50 to 70 mg/dL/h (2.8-3.9 mmol/L). If clinical circumstances prevent use of insulin infusion, then the insulin can be given intramuscularly. An initial 0.1 U/kg of regular insulin is given intravenously, and at the same time, the same size dose is given intramuscularly. Subsequently, regular insulin is given intramuscularly hourly at a dose of 0.1 U/kg until the blood glucose falls to around 250 mg/dL, when the insulin can be given subcutaneously. Insulin therapy, either as a continuous infusion or as injections given every 1 to 2 hours, should be continued until arterial pH has normalized. Patients who normally take long acting basal insulins (insulin glargine or insulin detemir or insulin degludec) can be given their usual maintenance doses during initial treatment of their diabetic ketoacidosis. The continuation of their subcutaneous basal insulins means that lower doses of intravenous insulin will be needed, and there will be a smoother transition from intravenous insulin infusion to the subcutaneous regimen.

3. **Potassium replacement.** Total body potassium loss from polyuria and vomiting may be as high as 200 mEq. However, because of shifts of potassium from cells into the extracellular space as a consequence of acidosis, serum potassium is usually normal to slightly elevated prior to institution of treatment. As the acidosis is corrected, potassium flows back into the cells, and hypokalemia can develop if potassium replacement is not instituted. If the patient is not uremic and has an adequate urine output, potassium chloride in doses of 10 to 30 mEq/h should be infused during the second and third hours after beginning therapy. Replacement should be started sooner, if the initial serum potassium is inappropriately normal or low, and should be delayed, if serum potassium fails to respond to initial therapy and remains above 5 mEq/L, as in cases of renal insufficiency. Occasionally, a patient may present with a serum potassium level less than 3.5 mEq/L, in which case insulin therapy should be delayed until the potassium level is corrected to greater than 3.5 mEq/L. An electrocardiogram can be help in monitoring the patient's potassium status: high peaked T waves are a sign of hyperkalemia, and flattened T waves with U waves are a sign of hypokalemia. Foods high in potassium content should be prescribed when the patient has recovered sufficiently to take food orally. Tomato juice has 14 mEq of potassium per 240 mL, and a medium-sized banana has about 10 mEq. Cooperative patients with only mild ketoacidosis may receive part or all of their potassium replacement orally.

4. **Sodium bicarbonate treatment.** The use of sodium bicarbonate in management of diabetic ketoacidosis has been questioned because clinical benefit was not demonstrated in one prospective randomized trial and because of the following potentially harmful consequences: (1) development of hypokalemia from rapid shift of potassium into cells if the acidosis is overcorrected, (2) tissue anoxia from reduced dissociation of oxygen from hemoglobin when acidosis is rapidly reversed (leftward shift of the oxygen dissociation curve), and (3) cerebral acidosis resulting from lowering of cerebrospinal fluid pH. It must be emphasized, however, that these considerations are less important when severe acidosis exists. It is therefore recommended that bicarbonate be administered to diabetic patients in ketoacidosis if the arterial blood pH is 7.0 or less with careful monitoring to prevent overcorrection.

One to two ampules of sodium bicarbonate (one ampule contains 44 mEq/50 mL) should be added to 1 L of 0.45% saline with 20 mEq KCl or to 400 mL of sterile water with 20 mEq KCl and infused over 1 to 2 hours. (**Note:** Addition of sodium bicarbonate to 0.9% saline would produce a markedly hypertonic solution that could aggravate the hyperosmolar state already present.) It can be repeated until the arterial pH reaches 7.1, but *it should not be given if the pH is 7.1 or greater*, because additional bicarbonate increases the risk of rebound metabolic alkalosis as ketones are metabolized. Alkalosis shifts potassium from serum into cells, which can precipitate a fatal cardiac arrhythmia. As noted earlier, serious consideration should be given to placement of a central venous catheter when administering fluids to severely ill patients with cardiovascular compromise.

5. **Phosphate.** Phosphate replacement is seldom required in treating diabetic ketoacidosis. However, if severe hypophosphatemia of less than 1 mg/dL (<0.35 mmol/L) develops during insulin therapy, a small amount of phosphate can be replaced per hour as the potassium salt. Correction of hypophosphatemia helps restore the buffering capacity of the plasma, thereby facilitating renal excretion of hydrogen. It also corrects the impaired oxygen dissociation from hemoglobin by regenerating 2,3-diphosphoglycerate. However, three randomized studies in which phosphate was replaced in only half of a group of patients with diabetic ketoacidosis did not show any apparent clinical benefit from phosphate administration. Moreover, attempts to use potassium phosphate as the sole means of replacing potassium have led to a number of reported cases of severe hypocalcemia with tetany. To minimize the risk of inducing tetany from too rapid replacement of phosphate, the average deficit of 40 to 50 mmol of phosphate should be replaced intravenously at a rate *no greater than 3 to 4 mmol/h* in a 60- to 70-kg person. A stock solution (Abbott) provides a mixture of 1.12 g KH_2PO_4 and 1.18 g K_2HPO_4 in a 5-mL single-dose vial (this equals 22 mmol of potassium and 15 mmol of phosphate). One-half of this vial (2.5 mL) should be added to 1 L of either 0.45% saline or 5% dextrose in water. Two liters of this solution, infused at a rate of 400 mL/h, corrects the phosphate deficit at the optimal rate of 3 mmol/h and provides 4.4 mEq of potassium per hour. Additional potassium should be administered as potassium chloride to provide a total of 10 to 30 mEq of potassium per hour, as noted earlier. If the serum phosphate remains below 2.5 mg/dL after this infusion, a repeat 5-hour infusion can be given.

6. **Hyperchloremic acidosis during therapy.** Because of the considerable loss of keto acids in the urine during the initial phase of therapy, substrate for subsequent regeneration of bicarbonate is lost, and correction of the total bicarbonate deficit is hampered. A portion of the bicarbonate deficit is replaced with chloride ions infused in large amounts as saline to correct the dehydration. In most patients, as the ketoacidosis clears during insulin replacement, a hyperchloremic, low-bicarbonate pattern emerges with a normal anion gap. This is a relatively

benign condition that reverses itself over the subsequent 12 to 24 hours once intravenous saline is no longer being administered. Using a balanced electrolyte solution with a pH of 7.4 and 98 mEq/L chloride instead of normal saline (pH ~5.5; chloride 154 mEq/L) has been reported to prevent the hyperchloremic acidosis.

TRANSITION TO SUBCUTANEOUS INSULIN REGIMEN

Once the diabetic ketoacidosis is controlled and the patient is awake and able to eat, subcutaneous insulin therapy can be initiated. Initially, the patient may still have significant tissue insulin resistance and may require a total daily insulin dose of ~0.6 U/kg. Half of the total daily dose can be given as long-acting basal insulin and the other half as short-acting insulin premeals. The patient should get injection of the basal insulin and a dose of the rapid-acting insulin analog with the first meal and the insulin infusion discontinued an hour later. The overlap of the subcutaneous insulin action and insulin infusion is necessary to prevent relapse of diabetic ketoacidiosis. In patients with preexisting diabetes, giving their basal insulin by subcutaneous injection at initiation of treatment of diabetic ketoacidosis simplifies the transition from intravenous to subcutaneous regimen. The increased tissue insulin resistance is only present for a few days at most and as the patient improves the doses of both basal and bolus insulins should be reduced to avoid hypoglycemia. In fact, a patient with new diagnosis of type 1 diabetes who still has significant β cell function may not require any basal insulin and only very low doses of rapid-acting insulin analogs before meals after full recovery from the ketoacidosis. Patients with type 2 diabetes and diabetes ketoacidosis due to severe illness may initially require insulin therapy but can often transition back to oral agents during outpatient follow-up.

Complications and Prognosis

Low-dose insulin infusion and fluid and electrolyte replacement combined with careful monitoring of patients' clinical and laboratory responses to therapy have dramatically reduced the mortality rates of diabetic ketoacidosis to less than 5%. However, this complication remains a significant risk in the aged who have mortality rates over 20% and in patients in profound coma in whom treatment has been delayed. Acute myocardial infarction and infarction of the bowel following prolonged hypotension worsen the outlook. Prior kidney dysfunction worsens prognosis because the kidney plays a key role in compensating for pH and electrolyte abnormalities. Symptomatic cerebral edema occurs primarily in the pediatric population. Risk factors for development include severe baseline acidosis, rapid correction of hyperglycemia, and excess volume administration in the first 4 hours. Onset of headache or deterioration in mental status during treatment should lead to consideration of this complication. Intravenous mannitol at a dosage of 1 to 2 g/kg given over 15 minutes is the mainstay of therapy. Excess crystalloid infusion can precipitate pulmonary edema. Acute respiratory distress syndrome is a rare complication of treatment for DKA.

Disposition

After recovery and stabilization, patients should receive intensive detailed instructions about how to avoid this potentially disastrous complication of diabetes mellitus. They should be taught to recognize the early symptoms and signs of ketoacidosis.

Urine ketones or capillary blood β-hydroxybutyrate should be measured in patients with signs of infection or in those using an insulin pump when capillary blood glucose is unexpectedly and persistently high. When heavy ketonuria and glycosuria persist on several successive examinations, supplemental regular insulin should be administered, and liquid foods such as lightly salted tomato juice and broth should be ingested to replenish fluids and electrolytes. Patients should be instructed to contact the physician if ketonuria persists and, especially, if vomiting develops, or if appropriate adjustment of the infusion rate on an insulin pump does not correct the hyperglycemia and ketonuria. Table 17–20 summarizes the guidelines for patients regarding ketone testing and what to do with the results. Recurrent episodes of severe diabetic ketoacidosis often indicate poor compliance with the insulin regimen, and these patients should receive intensive counseling.

2. HYPERGLYCEMIC, HYPEROSMOLAR STATE

This form of hyperglycemic coma is characterized by severe hyperglycemia, hyperosmolality, and dehydration in the absence of significant ketosis. It occurs in patients with mild or occult diabetes and patients are typically middle-aged or elderly. Lethargy and confusion develop as serum osmolality exceeds 300 mOsm/kg, and coma can occur if osmolality exceeds 330 mOsm/kg. Underlying renal insufficiency or congestive heart failure is common, and the presence of either worsens the prognosis. A precipitating event such as pneumonia, cerebrovascular accident, myocardial infarction, burns, or recent operation can often be identified. Certain drugs, such as phenytoin, diazoxide, glucocorticoids, and thiazide diuretics, have been implicated in its development, as have procedures associated with glucose loading such as peritoneal dialysis.

TABLE 17–20   Guidelines for treatment of ketones in patients with type 1 diabetes.

Check Ketones if Blood Glucose Persistently over 250 mg/dL or if Nauseated or Vomiting	
Blood ketones <1.5 mmol/L or urine ketones absent or small	Blood ketones >1.5 or urine ketones moderate or large
Drink plenty of fluids Give fast-acting insulin by syringe Monitor glucose levels	Call medical team for advise or go to urgent care or emergency room Drink plenty of fluids Give fast-acting insulin by syringe Monitor ketone levels

Note: Blood ketones <0.6 mmol/L are normal.

Pathogenesis

A partial or relative insulin deficiency may initiate the syndrome by reducing glucose utilization by muscle, fat, and liver, while promoting hyperglucagonemia and increasing hepatic glucose output. The result is hyperglycemia that leads to glycosuria and osmotic diuresis with obligatory water loss. The presence of even small amounts of insulin is believed to prevent the development of ketosis by inhibiting lipolysis in the adipose stores. Therefore, even though a low insulin-glucagon ratio promotes ketogenesis in the liver, the limited availability of precursor free fatty acids from the periphery restricts the rate at which ketones are formed. If a patient is unable to maintain adequate fluid intake because of an associated acute or chronic illness or has suffered excessive fluid loss (eg, from burns or therapy with diuretics), marked dehydration results. As plasma volume contracts, renal insufficiency develops; this, then, limits renal glucose excretion and contributes markedly to the rise in serum glucose and osmolality. As serum osmolality exceeds 320 to 330 mOsm/kg, water is drawn out of cerebral neurons, resulting in mental obtundation and coma.

Clinical Features

A. Symptoms and signs   The onset of the hyperglycemic, hyperosmolar, nonketotic state may be insidious, preceded for days or weeks by symptoms of weakness, polyuria, and polydipsia. A history of reduced fluid intake is common, whether due to inappropriate absence of thirst, gastrointestinal upset, or, in the case of elderly or bedridden patients, lack of access to water. A history of ingestion of large quantities of sugar-containing fluids, such as soft drinks or orange juice, can occasionally be obtained; these patients are usually less hyperosmolar than those in whom fluid intake was restricted. The absence of toxic features of ketoacidosis may retard recognition of the syndrome and thus delay institution of therapy until dehydration is profound. Because of this delay in diagnosis, the hyperglycemia, hyperosmolality, and dehydration in hyperglycemic, hyperosmolar, nonketotic coma is often more severe than in diabetic ketoacidosis.

Physical examination reveals the presence of profound dehydration (orthostatic fall in blood pressure and rise in pulse, supine tachycardia, or even frank shock, dry mucous membranes, decreased skin turgor). The patient may be lethargic, confused, or comatose. Kussmaul respirations are absent unless the precipitating event for the hyperosmolar state has also led to the development of metabolic acidosis (eg, sepsis or myocardial infarction with shock).

B. Laboratory findings   Severe hyperglycemia is present, with blood glucose values ranging from 800 to as high as 2400 mg/dL (44.4-133.2 mmol/L). In mild cases, where dehydration is less severe, dilutional hyponatremia as well as urinary sodium losses may reduce serum sodium to about 120 to 125 mEq/L—this protects, to some extent, against extreme hyperosmolality. Once dehydration progresses further, however, serum sodium can exceed 140 mEq/L, producing serum osmolalities of 330 to 440 mOsm/kg (normal, 280-295 mOsm/kg; see the section Diabetic Ketoacidosis for a convenient method for estimating serum

osmolality). Ketosis is usually absent or mild; however, a small degree of ketonuria may be present if the patient has not been eating because of illness. Acidosis is not a part of the hyperglycemic, hyperosmolar state, but it may be present (often lactic acidosis) because of other acute underlying conditions (eg, sepsis, acute renal failure, myocardial infarction). Prerenal azotemia is the rule with blood urea nitrogen frequently over 100 mg/dL.

The physician must initiate a careful search for the event that precipitated the episode of hyperglycemic, hyperosmolar state if it is not obvious after the initial history and physical examination. Chest x-rays and cultures of blood, urine, and other body fluids should be obtained to look for occult sources of sepsis. Cardiac enzymes and serial electrocardiograms can be ordered to look for evidence of silent myocardial infarction.

Treatment

There are some differences in fluid, insulin, and electrolyte replacement in this disorder, as compared with diabetic ketoacidosis. However, in common with the treatment of ketoacidotic patients, careful monitoring of the patient's clinical and laboratory response to therapy is essential.

1. **Fluid replacement.** The fluid deficit may be as much as 100 to 200 mL/kg or about 9 L. If circulatory collapse is present, fluid therapy should be initiated with isotonic saline. In all other cases, initial replacement with hypotonic (usually 0.45%) saline is preferable, because these patients are hyperosmolar with considerable loss of body water and excess solute in the vascular compartment. As much as 4 to 6 L of fluid may be required in the first 8 to 10 hours. Careful monitoring of fluid quantity and type, urine output, blood pressure, and pulse is essential. Placement of a central venous pressure catheter should be strongly considered to guide replacement of fluid, especially if the patient is elderly or has underlying renal or cardiac disease. Because insulin therapy decreases plasma glucose and therefore serum osmolality, a change to isotonic saline may be necessary at some time during treatment. Once blood glucose reaches 250 mg/dL, 5% dextrose in 0.45% or 0.9% saline solution should be substituted for the sugar-free fluids. The rate of dextrose infusion should be adjusted to maintain glycemic levels of 250 to 300 mg/dL (13.9-16.7 mmol/L) in order to reduce the risk of cerebral edema. An important endpoint of fluid therapy is to restore urinary output to 50 mL/h or more. When consciousness returns, oral fluids should be encouraged.

2. **Electrolyte replacement.** Hyperkalemia is less marked, and much less potassium is lost in the urine during the osmotic diuresis of hyperglycemic, hyperosmolar, nonketotic coma than in diabetic ketoacidosis. There is, therefore, less severe total potassium depletion, and less potassium replacement is needed to restore potassium stores to normal. However, because the initial serum potassium usually is not elevated and because it declines rapidly as insulin therapy allows glucose and potassium to enter cells, it is recommended that potassium replacement be initiated earlier than in ketotic patients: 10 mEq of potassium chloride can be added to the *initial* liter of fluid administered if the initial serum potassium is not elevated and if the patient is making urine. When serum phosphate falls below 1 mg/dL during insulin therapy, phosphate replacement can be given intravenously with the same precautions as those

outlined for ketoacidotic patients (see earlier). If the patient is awake and cooperative, part or all of the potassium and phosphate replacement can be given orally.

3. **Insulin therapy.** In general, less insulin is required to reduce the hyperglycemia of nonketotic patients than is the case for patients in diabetic ketoacidosis. In fact, fluid replacement alone can decrease glucose levels considerably. Insulin treatment should therefore be delayed unless the patient has significant ketonemia (β-hydroxybutyrate >1 mmol/L). Start the insulin infusion rate at 0.05 U/kg/h (no need to give bolus) and titrate to lower blood glucose by 50 to 70 mg/dL/h. Once the patient has stabilized and the blood glucose falls to around 250 mg/dL, insulin can be given subcutaneously.

Complications and Prognosis

The severe dehydration and low output state may predispose the patient to complications such as myocardial infarction, stroke, pulmonary embolism, mesenteric vein thrombosis, and disseminated intravascular coagulation. Fluid resuscitation remains the primary approach to the prevention of these complications. Low-dose heparin prophylaxis is reasonable but benefits of routine anticoagulation remain doubtful. Rhabdomyolysis is a recognized complication of the hyperosmolar state, and it should be looked for and treated.

The overall mortality rate of hyperglycemic, hyperosmolar, nonketotic coma is over 10 times that of diabetic ketoacidosis, chiefly because of its higher incidence in older patients, who may have compromised cardiovascular systems or associated major illnesses. When patients are matched for age, the prognoses of these two forms of hyperosmolar coma are reasonably comparable. When prompt therapy is instituted, the mortality rate can be reduced from nearly 50% to that related to the severity of coexistent disorders. After the patient is stabilized, the appropriate form of long-term management of the diabetes must be determined. Insulin treatment should be continued for a few weeks, but the patients usually recover sufficient endogenous insulin secretion to make a trial of diet or diet plus oral agents worthwhile. When the episode occurs in a patient who has known diabetes, then education of the patient and caregivers should be instituted. They should be taught how to recognize situations (gastrointestinal upset, infection) that predispose to recurrence of hyperglycemic, hyperosmolar state as well as detailed information on how to prevent the escalating dehydration (small sips of sugar-free liquids, increase in usual hypoglycemic therapy, or early contact with the physician) that culminates in hyperosmolar coma.

3. LACTIC ACIDOSIS

When severely ill diabetic patients present with profound acidosis and an anion gap over 15 mEq/L but relatively low or undetectable levels of keto acids in plasma, the presence of excessive plasma lactate (>5 mmol/L) should be considered, especially if other causes of acidosis such as uremia are not present.

Pathogenesis

Lactic acid is the end product of the anaerobic metabolism of glucose. Normally, the principal sources of this acid are the erythrocytes (which lack the enzymes for aerobic oxidation), skeletal muscle, skin, and brain. The chief pathway for removal of lactic acid is by hepatic (and to some degree renal) uptake for conversion first to pyruvate and eventually back to glucose, a process that requires oxygen. Lactic acidosis occurs when excess lactic acid accumulates in the blood. This can be the result of overproduction (tissue hypoxia), deficient removal (hepatic failure), or both (circulatory collapse). Lactic acidosis is not uncommon in any severely ill patient suffering from cardiac decompensation, respiratory or hepatic failure, septicemia, or infarction of the bowel or extremities. Type A lactic acidosis is associated with tissue hypoxia from hypovolemia or endotoxic shock and need not be associated with hyperglycemia. Type B lactic acidosis is defined as that which occurs in the absence of clinical evidence for tissue hypoxia and is associated with diabetes per se or with biguanide therapy.

With the discontinuance of phenformin therapy in the United States, lactic acidosis in patients with diabetes mellitus has become uncommon, but it still must be considered in the acidotic diabetic patient if the patient is seriously ill, and especially if the patient is receiving metformin therapy as well. Most cases of metformin-associated lactic acidosis occur in patients in whom there were contraindications to the use of metformin, in particular renal failure.

Clinical Features

A. Symptoms and signs The main clinical features of lactic acidosis are marked hyperventilation and mental confusion, which may progress to stupor or coma. When lactic acidosis is secondary to tissue hypoxia or vascular collapse, the clinical presentation is variable, reflecting that of the prevailing catastrophic illness. In the rare instance of idiopathic or spontaneous lactic acidosis, the onset is rapid (usually over a few hours), the cardiopulmonary status is stable, and mentation may be relatively normal.

B. Laboratory findings Plasma glucose can be low, normal, or high in diabetic patients with lactic acidosis, but usually it is moderately elevated. Plasma bicarbonate and arterial pH are quite low. An anion gap is present (calculated by subtracting the sum of the plasma bicarbonate and chloride from the plasma sodium; normal is 12-15 mEq/L). Ketones are usually absent from plasma, but small amounts may be present in urine if the patient has not been eating recently. Other causes of anion gap metabolic acidosis should be excluded—for example, uremia, diabetic or alcoholic ketoacidosis, and salicylate, methanol, ethylene glycol, or paraldehyde intoxication. In the absence of azotemia, hyperphosphatemia may be a clue to the presence of lactic acidosis.

The diagnosis is confirmed by demonstrating, in a sample of blood that is promptly chilled and separated, a plasma lactate concentration of 6 mmol/L or higher (normal is ~1 mmol/L). Failure to rapidly chill the sample and separate the plasma can lead to falsely high plasma lactate values as a result of continued glycolysis by the red blood cells. Frozen plasma remains stable for subsequent assay.

Treatment

The cornerstone of therapy is aggressive treatment of the precipitating cause. An adequate airway and good oxygenation should

be ensured. If hypotension is present, fluids and, if appropriate, pressor agents must be given to restore tissue perfusion. Appropriate cultures and empiric antibiotic coverage should be instituted in any seriously ill patient with lactic acidosis in whom the cause is not immediately apparent. Alkalinization with intravenous sodium bicarbonate to keep the pH above 7.2 has been recommended in the emergency treatment of severe lactic acidosis. However, there is no evidence that the mortality rate is favorably affected by administering bicarbonate, and the matter is at present controversial, particularly because of the hazards associated with bicarbonate therapy. Hemodialysis may be useful in those cases associated with metformin toxicity. Dichloroacetate, an anion that facilitates pyruvate removal by activating pyruvate dehydrogenase, reverses certain types of lactic acidosis in animals, but in a prospective controlled clinical trial involving 252 patients with lactic acidosis, dichloroacetate failed to alter either hemodynamics or survival.

CHRONIC COMPLICATIONS OF DIABETES MELLITUS (TABLE 17–21)

In most patients with diabetes, a number of pathologic changes occur at variable intervals during the course of the disease. These changes involve the vascular system for the most part; however, they also occur in the nerves, the skin, and the lens. In addition to these complications, patients with diabetes have an increased incidence of certain types of infections and may handle their infections less well than the general population.

Classifications of Diabetic Vascular Disease

Diabetic vascular disease is conveniently divided into two main categories: microvascular disease and macrovascular disease.

A. Microvascular disease   Disease of the smallest blood vessels, the capillary and the precapillary arterioles, is manifested mainly by thickening of the capillary basement membrane. Microvascular disease involving the retina leads to diabetic retinopathy, and disease involving the kidney causes diabetic nephropathy. Small vessel disease may also involve the heart, and cardiomegaly with heart failure has been described in diabetic patients with patent coronary arteries.

B. Macrovascular disease   Large vessel disease in diabetes is essentially an accelerated form of atherosclerosis. It accounts for the increased incidence of myocardial infarction, stroke, and peripheral gangrene in diabetic patients. Just as in the case of atherosclerosis in the general population, the exact cause of accelerated atherosclerosis in the diabetic population remains unclear. Abnormalities in vessel walls, platelets and other components of the clotting system, red blood cells, and lipid metabolism have all been postulated to play a role. In addition, there is evidence that coexistent risk factors, such as cigarette smoking and hypertension, may be important in determining the course of the disease.

TABLE 17–21   Chronic complications of diabetes mellitus.

Eyes
    Diabetic retinopathy
        Nonproliferative (background)
        Proliferative
    Cataracts
        Subcapsular (snowflake)
        Nuclear (senile)
Kidneys
    Intercapillary glomerulosclerosis
        Diffuse
        Nodular
    Infection
        Pyelonephritis
        Perinephric abscess
        Renal papillary necrosis
    Renal tubular necrosis
        Following dye studies (urograms, arteriograms)
Nervous System
    Peripheral neuropathy
        Distal, symmetric sensory loss
    Motor neuropathy
        Foot drop, wrist drop
        Cranial nerves III, IV, VI, VII
        Diabetic amyotrophy
    Autonomic neuropathy
        Postural hypotension
        Resting tachycardia
        Loss of sweating
        Gastroparesis
        Diabetic diarrhea
        Urinary bladder atony
        Impotence (may also be secondary to pelvic vascular disease)
Skin
    Diabetic dermopathy (shin spots)
    Necrobiosis lipoidica diabeticorum
    Candidiasis
    Foot and leg ulcers
        Neurotropic
        Ischemic
Cardiovascular System
    Heart disease
        Myocardial infarction
        Cardiomyopathy
    Peripheral vascular disease
        Ischemic ulcers: gangrene
    Cerebrovascular disease
Bones and Joints
        Diabetic cheiroarthropathy
        Dupuytren contracture
        Charcot joint
        Osteomyelitis
Unusual Infections
        Necrotizing fasciitis
        Necrotizing myositis
        Mucor meningitis
        Emphysematous cholecystitis
        Malignant otitis externa

Prevalence of Chronic Complications by Type of Diabetes

Although all of the known complications of diabetes can be found in both types of the disease, some are more common in one type than in the other. In type 1 diabetes, end-stage renal disease

develops in up to 40% of patients, compared with less than 20% of patients with type 2 diabetes. Although blindness occurs in both types, it occurs more commonly as a result of severe proliferative retinopathy, vitreous hemorrhages, and retinal detachment in type 1 disease, whereas macular edema and ischemia are the usual cause in type 2 disease. Similarly, although diabetic neuropathy is common in both type 1 and type 2 diabetes, severe autonomic neuropathy with gastroparesis, diabetic diarrhea, resting tachycardia, and postural hypotension is much more common in type 1.

In patients with type 1 diabetes, complications from end-stage renal disease are a major cause of death, whereas patients with type 2 diabetes are more likely to have macrovascular diseases leading to myocardial infarction and stroke as the main causes of death. Cigarette use adds significantly to the risk of both microvascular and macrovascular complications in diabetic patients.

Molecular Mechanisms by Which Hyperglycemia Causes Microvascular and Macrovascular Damage

Epidemiological data and prospective intervention studies such as the DCCT have confirmed the central role of glucose in the development of chronic diabetic complications.

Four molecular mechanisms of glucose-induced damage have been proposed: (1) increased polyol pathway flux; (2) increased intracellular advanced glycation end product (AGE) formation; (3) activation of protein kinase C; and (4) increased hexosamine pathway flux.

Increase flux through the polyol pathway consumes NADPH. Because this cofactor is needed to regenerate reduced glutathione, NADPH depletion is predicted to exacerbate intracellular oxidative stress and cause cellular injury. Inhibitors of aldose reductase, the first enzyme in the polyol pathway have been shown to improve diabetic neuropathy.

Intracellular autoxidation of glucose results in production of intracellular dicarbonyls (glyoxal, 3-deoxyglucosone, methylglyoxal), also referred to as AGE precursors. These precursors damage target tissues by modifying intracellular and extracellular proteins and matrix components. Intracellular protein modifications may alter cellular functions. Modifications of extracellular matrix proteins result in abnormal interactions with other matrix proteins and integrins. The modified plasma proteins bind to receptors on endothelial cells, mesangial cells, and macrophages causing expression of cytokines and growth factors including interleukin 1, IGF-1, TNF-α, transforming growth factor-β (TGF-β), macrophage colony stimulating factor, granulocyte-macrophage stimulating factor, platelet-derived growth factor, thrombomodulin, tissue factor, vascular cell adhesion molecule 1 (VCAM 1), and vascular endothelial growth factor (VEGF). Induction of VEGF has been implicated in the vascular hyperpermeability associated with diabetes.

Protein kinase C isoforms β and δ are activated by diacylglycerol whose levels are increased by elevated intracellular glucose. Activation of these isoforms leads to alterations in expression of endothelial nitric oxide synthase, endothelin 1, VEGF, TGF-β, PAI-1, and activation of nuclear factor κB and NADPH oxidases.

Inhibitors of the protein kinase β isoform improve retinopathy and nephropathy in experimental models.

Hyperglycemia increases hexosamine pathway flux by providing more fructose-6-phosphate for the rate-limiting enzyme of the pathway glutamine: fructose-6-phosphate amidotransferase. Activity of this pathway leads to increased donation of N-acetylglucosamine moieties to serine and threonine moieties of complication-promoting factors such as PAI-1 or TGF-β.

It has been proposed that all four of these pathways are associated with overproduction of superoxide by mitochondria. High ambient glucose leads to increased substrate flux through glycolysis and the tricarboxylic acid cycle. This leads to increased potential difference across the inner mitochondrial membrane and generation of superoxide by the electron transport chain. The increased production of superoxide reduces glyceraldehyde-3-phosphate dehydrogenase (GAPDH) activity, which in turn leads to upstream increase in intracellular glucose and accumulation of glycolytic intermediates such as glyceraldehyde-3-phosphate and fructose-6-phosphate. The increased intracellular glucose leads to increased flux through the polyol pathway and also is the primary initiating event in the formation of both intracellular and extracellular AGEs. The glycolytic intermediates are important initiators of the hexosamine pathway (fructose-6-phosphate) or the protein kinase C pathway (glyceraldehyde-3-phosphate).

Genetic Factors in Susceptibility to Development of Chronic Complications of Diabetes

Although no genetic susceptibility genes have been identified as yet, two unrelated observations indicate that approximately 40% of people with diabetes may be unusually susceptible to the ravages of hyperglycemia or other metabolic sequelae of an inadequate insulin effect.

(1) In one retrospective study of 164 juvenile-onset diabetic patients with a median age at onset of 9 years, 40% were incapacitated or dead from end-stage renal disease with proliferative retinopathy after a 25-year follow-up, whereas the remaining subjects were either mildly affected (40%) or had no clinically detected microvascular disease (20%). This study was completed long before the availability of glycemic self-monitoring methodology, so it is unlikely that any of these patients were near optimal glycemic control.

(2) Data from renal transplantation indicate that only about 40% of normal kidneys developed evidence of moderate to severe diabetic nephropathy within 6 to 14 years of being transplanted into diabetic subjects with end-stage renal failure, whereas as many as 60% were only minimally affected. These observations support the hypothesis that although approximately 60% of people suffer only minimal consequences from hyperglycemia and other metabolic hazards of insulin insufficiency, 40% or so suffer severe, potentially catastrophic microvascular complications if the disease is poorly controlled. The genetic mechanisms for this increased susceptibility are as yet

unknown but could be related to one or more of the molecular mechanisms outlined earlier. Identification of the genetic mechanism(s) would be very helpful in justifying more intensive insulin therapy in the group susceptible to complications in an effort to achieve near-normalization of blood glucose. The remaining 60% of less susceptible individuals might then be spared the inconveniences of strict glycemic control as well as the risks of hypoglycemia inherent in present methods of intensive insulin therapy.

SPECIFIC CHRONIC COMPLICATIONS OF DIABETES MELLITUS (TABLE 17–21)

1. OPHTHALMOLOGIC COMPLICATIONS

Diabetic Retinopathy

For early detection of diabetic retinopathy, adolescent or adult patients who have had type 1 diabetes for more than 5 years and *all* patients with type 2 diabetes should be referred to an ophthalmologist for examination and follow-up. In patients with type 1 diabetes, after 10 to 15 years, 25% to 50% of patients show some signs of retinopathy. This prevalence increases to 75% to 95% after 15 years and approaches 100% after 30 years of diabetes. In patients with type 2 diabetes, 60% have nonproliferative retinopathy after 16 years. When hypertension is present in a patient with diabetes, it should be treated vigorously because hypertension is associated with an increased incidence and accelerated progression of diabetic retinopathy.

A. Pathogenesis and clinical features   It is now recognized that diabetic retinopathy is not only a microvascular complication of endothelial dysfunction but also a neurodegenerative disease. There is thinning of the inner nuclear layer; reduction in synapse numbers and synaptic proteins and changes in dendritic morphology.

Two main categories of diabetic retinopathy exist: nonproliferative and proliferative. Diabetic macular edema can occur at any stage but more prevalent in the later phases.

Nonproliferative (background) retinopathy (Figure 17–15) represents the earliest stage of retinal involvement by diabetes and is characterized by such changes as microaneurysms, dot hemorrhages, exudates, and retinal edema. During this stage, the retinal capillaries leak proteins, lipids, or red cells into the retina. When this process occurs in the macula (clinically significant macular edema), the area of greatest concentration of visual cells, there is interference with visual acuity; this is the most common cause of visual impairment in type 2 diabetes.

Proliferative retinopathy (Figure 17–16) involves the growth of new capillaries and fibrous tissue within the retina and into the vitreous chamber. It is a consequence of small vessel occlusion, which causes retinal hypoxia; this in turn stimulates new vessel growth. New vessel formation may occur at the optic disc or elsewhere on the retina. Proliferative retinopathy can occur in both types of diabetes but is more common in type 1, developing about 7 to 10 years after onset of symptoms, with a prevalence of 25% after 15 years' duration. Prior to proliferation of new capillaries, a preproliferative phase often occurs in which arteriolar ischemia is manifested as cotton-wool spots (small infarcted areas of retina).
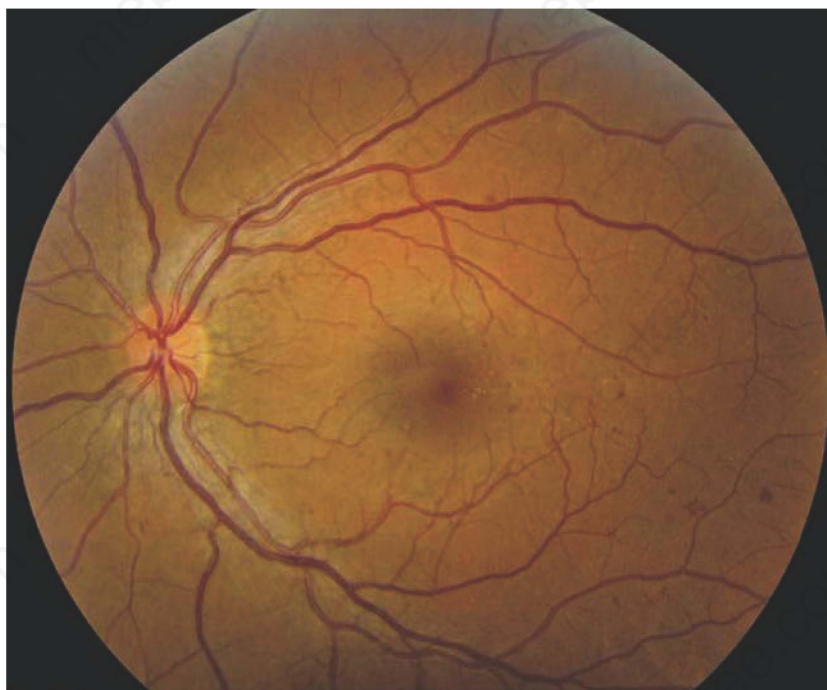
FIGURE 17–15 Nonproliferative diabetic retinopathy with intraretinal hemorrhages and microaneurysms along inferotemporal arcade; and hard exudates temporal to macula. (Used with permission from Dr. Jacque Duncan, University of California, San Francisco.)
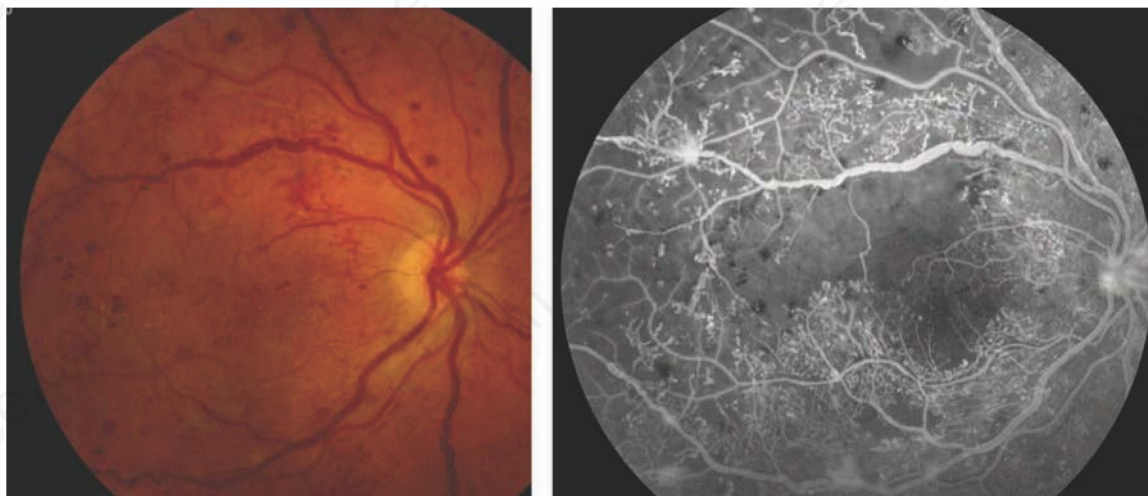
FIGURE 17–16 Proliferative diabetic retinopathy with neovascularization of the disc, venous beading, hemorrhages, and cotton wool spots nasal to the optic nerve. Early frame of fluorescein angiogram shows extensive macular capillary nonperfusion and early leakage from neovascularization along the superotemporal arcade. (Used with permission from Dr. Jacque Duncan, University of California, San Francisco.)

Vision is usually normal until vitreous hemorrhage or retinal detachment occurs. Proliferative retinopathy is a leading cause of blindness in the United States, particularly because it increases the risk of retinal detachment.

B. Treatment Extensive scatter xenon or argon photocoagulation and focal treatment of new vessels reduce severe visual loss in those cases in which proliferative retinopathy is associated with recent vitreous hemorrhages or in which extensive new vessels are located on or near the optic disk. Vitreoretinal surgery is used for sight threatening complications of proliferative diabetic retinopathy such as vitreous hemorrhage, tractional retinal detachment, and epimacular proliferation.

Macular edema, which is more common than proliferative retinopathy in patients with type 2 diabetes (up to 20% prevalence), has a guarded prognosis, but it has also responded to scatter therapy with improvement in visual acuity if detected early. Intravitreal anti-vascular endothelial growth factor (anti-VEGF) agents (ranibizumab, bevacizumab, pegaptanib, and aflibercept) have emerged as new treatments for diabetic macular edema and in clinical trials are better than laser therapy in preserving and improving vision. A significant proportion of patients, however, may still require focal or grid laser treatment. Intravitreal corticosteroids are effective and are used in patients affected by persistent or refractory diabetic macular edema, especially in pseudophakic eyes. Avoiding tobacco use and correction of associated hypertension are important therapeutic measures in the management of diabetic retinopathy. There is no contraindication to using aspirin in patients with proliferative retinopathy.

Cataracts

Two types of cataracts occur in diabetic patients: subcapsular and senile. **Subcapsular cataract** occurs predominantly in patients with type 1 diabetes, may come on fairly rapidly, and has a significant correlation with the hyperglycemia of uncontrolled diabetes. This type of cataract has a flocculent or *snowflake* appearance and develops just below the lens capsule.

Senile cataract represents a sclerotic change of the lens nucleus. It is by far the most common type of cataract found in either diabetic or nondiabetic adults and tends to occur at a younger age in diabetic patients, particularly when glycemic control is poor.

Two separate abnormalities found in diabetic patients, both of which are related to elevated blood glucose levels, may contribute to the formation of cataracts: (1) glycosylation of the lens protein, and (2) an excess of sorbitol, which is formed from the increased quantities of glucose found in the insulin-independent lens. Accumulation of sorbitol leads to osmotic changes in the lens that ultimately result in fibrosis and cataract formation.

Glaucoma

Glaucoma occurs in approximately 6% of persons with diabetes. It is generally responsive to the usual therapy for open-angle disease. Closed-angle glaucoma can result from neovascularization of the iris in diabetic patients, but this is relatively uncommon except after cataract extraction when accelerated new vessel growth involving the angle of the iris obstructs outflow.

2. RENAL COMPLICATIONS

Diabetic Nephropathy

A. Pathogenesis and clinical findings About 4000 cases of end-stage renal disease due to diabetic nephropathy occur annually among diabetic patients in the United States. This represents about one-third of all patients being treated for renal failure.

The cumulative incidence of nephropathy differs between the two major types of diabetes. Patients with type 1 diabetes, who have not received intensive insulin therapy and have had only fair to poor glycemic control, have a 30% to 40% chance of having nephropathy after 20 years—in contrast to the much lower frequency in patients with type 2 diabetes, who are not receiving intensive therapy, in whom only about 15% to 20% develop clinical renal disease. However, because so many more individuals are affected with type 2 diabetes, end-stage renal disease is much more prevalent in people with type 2 diabetes in the United States and especially throughout the rest of the world. There is no question that improved glycemic control and more effective therapeutic measures to correct hypertension can reduce the incidence of end-stage renal disease in both types of diabetes in the future.

Diabetic nephropathy is initially manifested by proteinuria; subsequently, as kidney function declines, urea and creatinine accumulate in the blood. Thickening of capillary basement membranes and of the mesangium of renal glomeruli produces varying degrees of glomerulosclerosis and renal insufficiency. Diffuse glomerulosclerosis is more common than nodular intercapillary glomerulosclerosis (Kimmelstiel-Wilson lesions); both produce heavy proteinuria.

Sensitive radioimmunoassay methods have permitted detection of microgram concentrations of urinary albumin. Conventional 24-hour urine collections, in addition to being inconvenient for patients, also show wide variability of albumin excretion, since several factors such as sustained erect posture, dietary protein, and exercise tend to increase albumin excretion rates. For these reasons, an albumin-creatinine ratio in an early morning spot urine collected upon awakening is preferable. In the early morning spot urine, a ratio of albumin (μg/L) to creatinine (mg/L) of less than 30 μg/mg creatinine is normal, and a ratio of 30 to 300 μg/mg creatinine suggests abnormal microalbuminuria. At least two early morning spot urine collections over a 3- to 6-month period should be abnormal before a diagnosis of microalbuminuria is made.

Subsequent renal failure can be predicted by persistent urinary albumin excretion rates exceeding 30 μg/mg creatinine. Glycemic control as well as a protein diet of 0.8 g protein/kg body weight/d may reduce both the hyperfiltration and the elevated microalbuminuria in patients in the early stages of diabetes and those with incipient diabetic nephropathy. Antihypertensive therapy also decreases microalbuminuria. Evidence from some studies—but not the UKPDS—supports a specific role for ACE inhibitors in reducing intraglomerular pressure in addition to lowering systemic blood pressure. An ACE inhibitor (captopril, 50 mg twice daily) in normotensive diabetics impedes progression to proteinuria and prevents the increase in albumin excretion rate. Since microalbuminuria has been shown to correlate with elevated nocturnal systolic blood pressure, it is possible that normotensive diabetic patients with microalbuminuria have slightly elevated systolic blood pressure during sleep, which is lowered during antihypertensive therapy. This action may contribute to the reported efficacy of ACE inhibitors in reducing microalbuminuria in normotensive patients.

If treatment is inadequate, then the disease progresses with proteinuria of varying severity, occasionally leading to nephrotic syndrome with hypoalbuminemia, edema, and an increase in circulating LDL cholesterol as well as progressive azotemia. In contrast to all other renal disorders, the proteinuria associated with diabetic nephropathy does not diminish with progressive renal failure (patients continue to excrete 10-11 g daily as creatinine clearance diminishes). As renal failure progresses, there is an elevation in the renal threshold at which glycosuria appears.

Patients with diabetic nephropathy should be evaluated and followed by a nephrologist. There has been gradual improvement in quality of life of diabetic patients receiving dialysis but mortality remains higher than in nondiabetic patients. During 5 years of follow-up in a European registry study, the mortality rate in people with diabetes receiving dialysis was 226.9 deaths/1000 patient years whereas the rate was 151.4 deaths/1000 patient years in people receiving dialysis who did not have diabetes.

Renal transplantation, especially from related donors, is often successful. For patients with compatible donors and no contraindications (such as severe cardiovascular disease), it is the treatment of choice. Diabetic nephropathy accounts for about 20% of kidney transplantations performed annually in the United States.

Necrotizing Papillitis

This unusual complication of pyelonephritis occurs primarily in diabetic patients. It is characterized by fever, flank pain, pyuria, and sloughing of renal papillae in the urine. It is treated by intravenous administration of appropriate antibiotics.

Renal Decompensation After Administration of Radiographic Dyes

The use of radiographic contrast agents in diabetic patients with reduced creatinine clearance has been associated with the development of acute renal failure. Contrast-induced nephropathy is defined as an increase in serum creatinine of at least 0.5 mg/dL or 25% compared with baseline after exposure to intravenous contrast. The increase in serum creatinine occurs shortly after the procedure and peaks 3 to 5 days later before improving over the next 1 to 3 weeks. Although frequently transient, it can cause permanent impairment of renal function. Diabetic patients with normal renal function do not appear to be at increased risk for contrast nephropathy. If a contrast study is considered essential, patients with a serum creatinine of 1.5 to 2.5 mg/dL should be adequately hydrated before the procedure. Hydration with saline has been the cornerstone of contrast nephropathy prevention: intravenous saline 1 mL/kg/h is started 12 hours before the procedure and continued for 12 hours afterward. Some studies suggest that sodium bicarbonate infusions are more effective, and an alternate option is to infuse sodium bicarbonate: 150 mL of sodium bicarbonate (1 mEq/mL) is added to 1 L of 5% dextrose and infused at 3.5 mL/kg/h for 1 hour before the procedure; 1.2 mg/kg/h during the procedure and for 6 hours afterward. *N*-acetylcysteine has also been shown in some trials to decrease the incidence of contrast nephropathy. One regimen consists of using oral *N*-acetylcysteine 600 mg twice a day starting the day before the procedure for a total of four doses. A combination of *N*-acetylcysteine and intravenous saline or intravenous sodium

bicarbonate may be more beneficial. Radiographic contrast material should not be given to a patient with a serum creatinine greater than 3 mg/dL unless the potential benefit outweighs the high risk of acute renal failure.

3. NEUROLOGIC COMPLICATIONS (DIABETIC NEUROPATHY)

Peripheral and autonomic neuropathies are the two most common complications of both types of diabetes. Up to 50% of patients with type 2 diabetes are affected. The pathogenesis of both types of neuropathy is poorly understood. Some lesions, such as the acute cranial nerve palsies and diabetic amyotrophy, have been attributed to ischemic infarction of the involved peripheral nerve. The much more common symmetric sensory and motor peripheral neuropathies and autonomic neuropathy are felt to be due to metabolic complications.

Unfortunately, there is no consistently effective treatment for any of the neuropathies. However, several long-term clinical trials have definitively shown that normalization of blood glucose levels can prevent development and progression of this devastating complication.

Peripheral Neuropathy

A. Distal symmetric polyneuropathy This is the most common form of diabetic peripheral neuropathy in which loss of function appears in a stocking-glove pattern and is due to an axonal neuropathic process. Longer nerves are especially vulnerable—hence the impact on the foot. Both motor and sensory nerve conduction is delayed in the peripheral nerves, and ankle jerks may be absent.

Sensory involvement usually occurs first and is generally bilateral, symmetric, and associated with dulled perception of vibration, and temperature. Pain, when present, can range from mild discomfort to severe incapacitating symptoms (discussed later). The sensory deficit may eventually be of sufficient degree to prevent patients from feeling pain. Patients with symptoms of sensory neuropathy should be examined with a 5.07 Semmes-Weinstein filament. Those who cannot feel the filament must be considered at risk for unperceived neuropathic injury.

The denervation of the small muscles of the foot results in clawing of the toes and displacement of the submetatarsal fat pads anteriorly. These changes, together with the joint and connective tissue changes, alter the biomechanics of the foot, increasing plantar pressures. This combination of decreased pain threshold, abnormally high foot pressures, and repetitive stress (eg, walking) can lead to calluses and ulcerations in the high-pressure areas such as over the metatarsal heads (Figure 17-17). Peripheral neuropathy also predisposes to the development of Charcot arthropathy. Other predisposing factors for this condition include autonomic neuropathy and trauma. The acute Charcot foot presents with pain and swelling, and if the condition is untreated, it leads to a rocker bottom deformity and ulceration. The early radiologic changes are of joint subluxation and periarticular fractures. As the

FIGURE 17–17 Neuropathic ulceration over first metatarsal head.

process progresses, there is frank osteoclastic destruction leading to deranged and unstable joints particularly in the midfoot.

Not surprisingly, the key issue for healing of neuropathic ulcers in a foot with good vascular status is mechanical unloading. Additionally, any infection should be treated with debridement and appropriate antibiotics. Healing times of 8 to 10 weeks are typical. In the occasional patient, where healing appears refractory, platelet-derived growth factor (becaplermin) should be considered for local application. A postmarket epidemiologic study showed increased cancer deaths in patients who had used three or more tubes of becaplermin on their leg or feet ulcers, and there is now a boxed warning in the drug label. Once ulcers are healed, therapeutic footwear is key to preventing recurrences. Custom-molded shoes are reserved for patients with significant foot deformities. Other patients with neuropathy may require accommodative insoles that distribute the load over as wide an area as possible. Patients with foot deformities and loss of protective threshold should seek regular care from a podiatrist. They should be educated about appropriate footwear, and those patients with loss of protective threshold should be instructed to inspect their feet daily for reddened areas, blisters, abrasions, or lacerations.

In some patients, hypersensitivity to light touch and occasionally severe burning pain, particularly at night, can become physically and emotionally disabling. Nortriptyline or desipramine in doses of 25 to 150 mg/d orally may provide dramatic relief for pain from diabetic neuropathy, often within 48 to 72 hours. This rapid response is in contrast to the 2 or 3 weeks required for an antidepressive effect. Patients often attribute the benefit to having a full night's sleep. Mild to moderate morning drowsiness is a side effect that generally improves with time or can be lessened by giving the medication several hours before bedtime. This medication should not be continued if improvement has not occurred after 5 days of therapy. Amitriptyline, 25 to 75 mg orally at bedtime can also be used but has more anticholinergic effects. Tricyclic antidepressants, in combination with the phenothiazine,

fluphenazine, have been shown in two studies to be efficacious in painful neuropathy, with benefits unrelated to relief of depression.

Gabapentin (900-1800 mg/d in three divided doses) has also been shown to be effective in the treatment of painful neuropathy and should be tried if the tricyclic drugs prove ineffective. Pregabalin, a congener of gabapentin, has been shown in an 8-week study to be more effective than placebo in treating painful diabetic peripheral neuropathy. However, this drug was not compared with an active control. Also, because of its abuse potential, it has been categorized as a Schedule V controlled substance. Duloxetine (60-120 mg), a serotonin and norepinephrine reuptake inhibitor, has been approved for the treatment of painful diabetic neuropathy. In clinical trials, this drug reduced the pain sensitivity score by 40% to 50%. Capsaicin, a topical irritant, has been found to be effective in reducing local nerve pain; it is dispensed as a cream (Zostrix 0.025% and Zostrix-HP 0.075%) to be rubbed into the skin over the painful region two to four times daily. Gloves should be used for application, because hand contamination could result in discomfort if the cream comes in contact with eyes or sensitive areas such as the genitalia. Five percent lidocaine patch applied over an area of maximal pain has been reported to be of benefit.

Rapid normalization of glucose levels in a patient with long standing poor control can precipitate an acute painful neuropathy (also called "insulin neuritis" or "treatment induced neuropathy"). Autonomic dysfunction may also be present. The symptoms improve with time. The initial report of this condition noted that stopping the insulin treatment resolved the symptoms. There is, however, no conclusive evidence that deliberately raising glucose levels will ameliorate the symptoms. Treatment induced neuropathy can probably be avoided by lowering glucose levels gradually.

Diabetic neuropathic cachexia is a syndrome characterized by a symmetric peripheral neuropathy associated with profound weight loss (up to 60% of total body weight) and painful dysesthesias affecting the proximal lower limbs, the hands, or the lower trunk. Treatment is usually with insulin and analgesics. The prognosis is usually good, and patients typically recover their baseline weight with resolution of the painful sensory symptoms within 1 year.

B. Isolated peripheral neuropathy   Involvement of the distribution of only one nerve (mononeuropathy) or of several nerves (mononeuropathy multiplex) is characterized by sudden onset with subsequent recovery of all or most of the function. This neuropathology has been attributed to vascular ischemia or traumatic damage. Cranial and femoral nerves are commonly involved, and motor abnormalities predominate. The patient with cranial nerve involvement usually presents with diplopia. Clinical examination reveals signs of single third, fourth, or sixth nerve weakness, and the pupil is spared. A full recovery of function occurs in 6 to 12 weeks. Diabetic amyotrophy presents with onset of severe pain in the front of the thigh. Within a few days or weeks of the onset of pain, the patient develops weakness and wasting of the quadriceps. Usually as the weakness appears, the pain tends to improve. Management includes analgesia and improved diabetic control. The symptoms improve over 6 to 18 months.

Autonomic Neuropathy

Neuropathy of the autonomic nervous system is common in patients with diabetes of long duration and can be a very disconcerting clinical problem. It can affect many diverse visceral functions including blood pressure and pulse, gastrointestinal activity, bladder function, and erectile function. Treatment is directed specifically at each abnormality.

Involvement of the gastrointestinal system may be manifested by nausea, vomiting, and postprandial fullness (from gastric atony); symptoms of reflux or dysphagia (from esophageal involvement); constipation and recurrent diarrhea, especially at night (from involvement of the small bowel and colon); and fecal incontinence (from anal sphincter dysfunction). Gall bladder function is altered, and this enhances stone formation.

Gastroparesis should be considered in type 1 diabetic patients in whom unexpected fluctuations and variability in their blood glucose levels develops after meals. Radiographic studies of the stomach and radioisotopic examination of gastric emptying after liquid and solid meals are of diagnostic value in these patients. Metoclopramide has been of some help in treating diabetic gastroparesis. It is a dopamine antagonist that has central antiemetic effects as well as a cholinergic action to facilitate gastric emptying. It is given in a dose of 10 mg orally three or four times a day, 30 minutes before meals and at bedtime. Drowsiness, restlessness, fatigue, and lassitude are common adverse effects. Tardive dyskinesia and extrapyramidal effects can occur especially when the drug is used for longer than 3 months and the FDA has cautioned against its chronic use. Domperidone is also a dopamine antagonist and as effective as metoclopramide but with less central side effects. The dose is 10 to 20 mg three or four times a day. It can cause prolongation of the QT interval and a baseline ECG and follow-up ECG should be performed. The drug is not licensed for sale in the United States but the FDA will allow its use through the "expanded access to investigational drugs program." Erythromycin appears to bind to motilin receptors in the stomach and has been found to improve gastric emptying in doses of 250 mg three times daily over the short term, but its effectiveness seems to diminish over time. Gastric electrical stimulation has been reported to improve symptoms and quality of life indices in patients with gastroparesis refractory to pharmacologic therapy.

Diarrhea associated with autonomic neuropathy has occasionally responded to broad-spectrum antibiotic therapy (such as rifaximin, metronidazole, amoxicillin/clavulanate, ciprofloxacin, or doxycycline), although it often undergoes spontaneous remission. Refractory diabetic diarrhea is often associated with impaired sphincter control and fecal incontinence. Therapy with loperamide, 4 to 8 mg daily, or diphenoxylate with atropine, two tablets up to four times a day, may provide relief. In more severe cases, tincture of paregoric or codeine (60-mg tablets) may be required to reduce the frequency of diarrhea and improve the consistency of the stools. Clonidine has been reported to lessen diabetic diarrhea; however, its usefulness is limited by its tendency to cause orthostatic hypotension in these patients who already have autonomic neuropathy. Constipation usually responds to stimulant laxatives such as senna. Metamucil and other bulk-providing

agents may relieve either the diarrhea or the constipation phases, which often alternate.

Inability to completely empty the bladder can sometimes occur. Bethanechol in doses of 10 to 50 mg three times a day has occasionally improved emptying of the atonic urinary bladder. Catheter decompression of the distended bladder has been reported to improve its function, and considerable benefit has been reported after surgical severing of the internal vesicle sphincter.

Use of Jobst fitted stockings, tilting the head of the bed, and arising slowly from the supine position are useful in minimizing symptoms of orthostatic hypotension. Some patients may require the addition of a mineralocorticoid such as fludrocortisone acetate (0.1-0.2 mg twice daily). Fludrocortisone therapy, however, can result in supine hypertension and hypokalemia. Midodrine (10 mg three times a day), an alpha adrenergic agonist, can be tried if Jobst stockings and sleeping upright prove ineffective in providing symptomatic relief.

Erectile dysfunction due to neuropathy differs from the psychogenic variety in that the latter may be intermittent (erections occur under special circumstances), whereas diabetic erectile dysfunction is usually persistent. To distinguish neuropathic or psychogenic erectile dysfunction from the erectile dysfunction caused by aortoiliac occlusive disease or vasculopathy, papaverine is injected into the corpus cavernosum of the penis. If the blood supply is competent, a penile erection occurs (see Chapter 12). Urinary incontinence, with large volumes of residual urine, and retrograde ejaculation can also result from pelvic neuropathy.

There are medical, mechanical, and surgical approaches available for treatment of erectile dysfunction. Penile erection depends on relaxation of the smooth muscle in the arteries of the corpus cavernosum, and this is mediated by nitric oxide-induced cyclic $3',5'$-guanosine monophosphate (cGMP) formation. cGMP-specific phosphodiesterase type 5 (PDE5) inhibitors impair the breakdown of cGMP and improve the ability to attain and maintain an erection. Sildenafil (Viagra), vardenafil (Levitra), and tadalafil (Cialis) have been shown in placebo-controlled clinical trials to improve erections in response to sexual stimulation. The recommended dose of sildenafil for most patients is one 50-mg tablet taken approximately 1 hour before sexual activity. The peak effect is at 1.5 to 2 hours, with some effect persisting for 4 hours. Patients with diabetes mellitus using sildenafil reported 50% to 60% improvement in erectile function. The maximum recommended dose is 100 mg. The recommended dose of both vardenafil and tadalafil is 10 mg. The doses may be increased to 20 mg or decreased to 5 mg based on efficacy and side effects. Tadalafil has been shown to improve erectile function for up to 36 hours after dosing. In clinical trials, only a few adverse effects have been reported—transient mild headache, flushing, dyspepsia, and, in some, altered color vision. Priapism can occur with these drugs and patients should be advised to seek immediate medical attention if an erection persists for longer than 4 hours. The PDE5 inhibitors potentiate the hypotensive effects of nitrates, and their use is contraindicated in patients who are concurrently using organic nitrates in any form. Caution is advised for men who have suffered a heart attack, stroke, or life-threatening arrhythmia within the previous 6 months; men who have resting hypotension or hypertension, and men who have a history of cardiac failure or have unstable angina. Rarely, a decrease in vision or permanent visual loss has been reported after PDE5 inhibitor use.

Intracorporeal injection of vasoactive drugs causes penile engorgement and erection. Drugs most commonly used include papaverine alone, papaverine with phentolamine, and alprostadil (prostaglandin E_1). Alprostadil injections are relatively painless, but careful instruction is essential to prevent local trauma, priapism, and fibrosis. Intraurethral pellets of alprostadil avoid the problem of injection of the drug.

External vacuum therapy (Erec-Aid System) is a nonsurgical treatment consisting of a suction chamber operated by a hand pump that creates a vacuum around the penis. This draws blood into the penis to produce an erection that is maintained by a specially designed tension ring inserted around the base of the penis and which can be kept in place for up to 20 to 30 minutes. Although this method is generally effective, its cumbersome nature limits its appeal.

Surgical implants of penile prostheses remain an option for those patients in whom the nonsurgical approaches are ineffective.

4. CARDIOVASCULAR COMPLICATIONS
Heart Disease

Microangiopathy occurs in the heart and may explain the existence of congestive cardiomyopathies found in diabetic patients without demonstrable coronary artery disease. Much more commonly, however, heart failure in patients with diabetes is a consequence of coronary atherosclerosis. Myocardial infarction is three to five times more common in diabetic patients than in age-matched controls and is the leading cause of death in patients with type 2 diabetes. Cardiovascular disease risk is increased in patients with type 1 diabetes as well, although the absolute risk is lower than in patients with type 2 diabetes. Women with diabetes lose the protection against myocardial infarction that is usually present during the childbearing years. The increased risk of atherosclerosis in people with diabetes may reflect the combination of hyperlipidemia, abnormalities of platelet adhesiveness, coagulation factors, hypertension, and oxidative stress and inflammation.

Large intervention studies of risk factor reduction in diabetes are lacking, but it is reasonable to assume that reducing these risk factors would have beneficial effects. Lowering LDL cholesterol reduces first events in patients without known coronary disease and secondary events in patients with known coronary disease. These intervention studies included some patients with diabetes, and the benefits of lowering LDL cholesterol were apparent in this group. The National Cholesterol Education Program clinical practice guidelines have designated diabetes as a coronary risk equivalent and have recommended that patients with diabetes should have an LDL cholesterol goal of less than 100 mg/dL. Lowering LDL cholesterol to 70 mg/dL may have additional benefit and is a reasonable target for most patients with type 2 diabetes who have multiple risk factors for cardiovascular disease. In people with diabetes, the association of blood pressure with microvascular and macrovascular events is a continuum. The UKPDS reported that a 10 mm Hg decrease in mean systolic pressure was associated with a risk reduction of 11% for myocardial infarction; 13% for

microvascular complication; and 15% for deaths related to diabetes. There was no threshold of risk for any endpoint and the lowest risk was in those patients whose systolic pressure was less than 120 mm Hg. Meta-analyses of clinical trials that included patients with diabetes report that if baseline systolic pressure is greater than 140, treatment to decrease blood pressure toward the normal range reduces risk of mortality and macrovascular and microvascular events. The data on treatment when baseline systolic pressure is less than 140 is conflicting. One meta-analysis found that treating patients with baseline systolic pressure less than 140 was associated with a reduced risk of stroke and albuminuria. Another meta-analysis, however, found that if baseline systolic blood pressure was less than 140, further treatment increased the risk of cardiovascular mortality. These meta-analyses included studies of diabetic patients with known cardiovascular disease, chronic renal failure, heart failure; and stroke and it is possible that including these very high-risk patients attenuated or even worsened outcomes from blood pressure treatment.

The antihypertensive regimen used does not seem to be that important. The key exception is heart failure where diuretics seem to beneficial. Dual renin-angiotensin blockade with a renin blocker, such as aliskiren, with ACE-inhibition or ARB blockade is not recommended because of increased risk of hyperkalema, hypotension and stroke. Similarly, combining an ARB and ACE inhibitor has also failed to show benefit and is associated with hypotension and renal dysfunction. Although evidence that ACE inhibitors or ARB are of particular benefit is sparse, these classes of drugs are well tolerated and are recommended as the first line antihypertensive treatment in patients with diabetes.

The ADA recommends lowering systolic blood pressure to less than 140 mm Hg and diastolic pressure to less than 90 mm Hg in people with diabetes. The systolic target of 130 mm Hg or less and diastolic of 80 mm Hg or less is recommended for the younger patient if it can be achieved without undue treatment burden. The Systolic Blood Pressure Intervention Trial (SPRINT) reported that treating to a systolic blood pressure of less than 120 mm Hg reduced cardiovascular events by 25% and death from cardiovascular causes by 43% during 3.26 years of follow-up. People with diabetes, however, were excluded from this study and it is unclear if the results are applicable to this population.

Aspirin (81-325 mg daily) inhibits thromboxane synthesis by platelets and is effective in reducing cardiovascular morbidity and mortality in patients who have a history of myocardial infarction or stroke (secondary prevention). It is unclear if aspirin prevents primary cardiovascular events in people with diabetes. The current recommendation is to give aspirin to those people with diabetes who are at increased risk for cardiovascular events (>10% 10-year risk of cardiovascular events). Typically this includes most 50-year-old men and 60-year-old women with one or more additional risk factors (smoking, hypertension, dyslipidemia, family history of premature cardiovascular disease, and albuminuria). Contraindications for aspirin therapy include age less than 21 years (because of risk of Reye syndrome), aspirin allergy, bleeding tendency (eg, anticoagulant therapy), recent gastrointestinal bleeding, or active hepatic disease. The Early Treatment Diabetic Retinopathy Study (ETDRS) showed that aspirin does not influence the course of proliferative retinopathy. There was no statistically significant difference in the severity of vitreous/preretinal hemorrhages or their rate of resolution between the aspirin and placebo groups. Thus, it appears that there is no contraindication to aspirin use to achieve cardiovascular benefit in diabetic patients who have proliferative retinopathy.

Peripheral Vascular Disease

Atherosclerosis is markedly accelerated in the larger arteries. It is often diffuse, with localized enhancement in certain areas of turbulent blood flow, such as at the bifurcation of the aorta or other large vessels. Clinical manifestations of peripheral vascular disease include ischemia of the lower extremities, impotence, and intestinal angina.

The incidence of **gangrene of the feet** in people with diabetes is 30 times that in age-matched controls. The factors responsible for the development of this condition, in addition to peripheral vascular disease, are small vessel disease, peripheral neuropathy with loss of both pain sensation and neurogenic inflammatory responses, and secondary infection. In two-thirds of patients with ischemic gangrene, pedal pulses are not palpable. In the remaining one-third who has palpable pulses, reduced blood flow through these vessels can be demonstrated by plethysmographic or Doppler ultrasound examination. Prevention of foot injury is imperative. Agents that reduce peripheral blood flow such as tobacco and propranolol should be avoided. Control of other risk factors such as hypertension is essential. Cholesterol-lowering agents are useful as adjunctive therapy when early ischemic signs are detected and when dyslipidemia is present. Patients should be advised to seek immediate medical care if a diabetic foot ulcer develops. Improvement in peripheral blood flow with endarterectomy and bypass operations is possible in certain patients.

5. SKIN AND MUCOUS MEMBRANE COMPLICATIONS

Chronic pyogenic infections of the skin may occur, especially in poorly controlled diabetic patients. Candidal infection can produce erythema and edema of intertriginous areas below the breasts, in the axillas, and between the fingers. It causes vulvovaginitis in most chronically uncontrolled diabetic women with persistent glucosuria and is a frequent cause of pruritus. While antifungal creams containing miconazole or clotrimazole offer immediate relief of vulvovaginitis, recurrence is frequent unless glucosuria is reduced.

In some patients with type 2 diabetes, poor glycemic control can cause a severe hypertriglycemia, which can present as eruptive cutaneous xanthomas and pancreatitis. The skin lesions appear as yellow morbilliform eruptions 2 to 5 mm in diameter with erythematous areolae. They occur on extensor surfaces (elbows, knees, buttocks) and disappear after triglyceride levels are reduced.

Necrobiosis lipoidica diabeticorum (Figure 17-18) are oval or irregularly shaped plaques with reddish demarcated borders and a glistening yellowish surface usually located over the anterior surfaces of the legs or the dorsal surfaces of the ankles. Rarely, lesions can occur on the hands, fingers, forearms, face, and scalp. The necrobiosis lesion usually starts out as an oval violaceous patch that slowly expands. The advancing border is red and the central

FIGURE 17–18 Necrobiosis lipoidica diabeticorum.

area turns yellow-brown. The thinning of the dermis in the center of the lesion leads to the shiny surface and prominent telangiectasia. It also allows the subcutaneous fat to become more visible hence the yellowish appearance. Pathologically, the lesions show degeneration of collagen, granulomatous inflammation of subcutaneous tissues and blood vessels, capillary basement membrane thickening, and obliteration of vessel lumina. The condition is associated with type 1 diabetes, although it can occur in patients with type 2 diabetes and also in people without diabetes. It occurs in about 0.3% of patients with diabetes, usually in patients in their 30s and 40s, and is about three times more common in women than in men. In some studies an association with microalbuminuria and retinopathy has been reported. Improving glycemic control may help the condition. The first-line treatment includes topical and subcutaneous corticosteroids. Second-line treatments include systemic steroids, cyclosporine, ticlodipine, nicotinamide, clofazimine, fumarate esters, intralesional etanercept, and topical psoralen with ultraviolet A radiation. Pulsed dye lasers can improve the appearance of telangiectasias. Flare-ups are frequent. No treatment is completely effective.

Shin spots are uncommon in adult diabetics. They are brownish, rounded, painless atrophic lesions of the skin in the pretibial area.

6. BONE AND JOINT COMPLICATIONS

Bone and joint complications are generally attributed to metabolic or vascular sequelae of long-standing diabetes.

Diabetic cheiroarthropathy is a syndrome of chronic progressive stiffness of the hand secondary to contracture and tightening of the skin over the joints. It is characterized by inability to flatten the palms against a flat surface (prayer sign). It is believed to be due to glycosylation of collagen and perhaps other proteins in connective tissue. It is associated with poor glycemic control and with longer duration of diabetes.

Dupuytren contractures consist of nodular thickening and contracture of the palmar fascia of the hand, producing flexure contractures of the fingers. It can occur in the absence of diabetes, but is more common in people with diabetes. The etiology is unclear but there may be an inflammatory component. Glucocorticoid injection into discrete nodules can sometimes help but the standard treatment is surgical fasciectomy.

Carpal tunnel syndrome occurs when the median nerve is compressed within the carpal tunnel. It is more common in people with diabetes, especially those who also have diabetic cheiroarthropathy. It is presumably due to glycosylation of collagen and other proteins in the connective tissue.

Patients with adhesive capsulitis of the shoulder (frozen shoulder) complain primarily of stiffness and loss of range of motion. They may also have shoulder pain. It occurs more frequently in people with long-standing diabetes—the incidence being two to four times higher than in the general population. The patients may also have diabetic cheiroarthropathy or Dupuytren contractures. In most cases it is a limited condition that responds to physical therapy. Patients should be warned that recovery may take 6 to 18 months. For a few patients surgery may be necessary. Adhesive capsulitis of the hip has also been described.

Data on bone mineral density and fracture risk in people with diabetes are contradictory. Bone mineral density has been reported to be low, normal, or high in type 2 diabetes patients. Type 2 diabetes patients do appear to be at increased risk for nonvertebral factures. Women with type 1 diabetes have an increased risk of fracture when compared to women without diabetes. It is likely that other factors such as duration of diabetes and diabetes complications such as neuropathy and renal disease affect both the bone mineral density and fracture risk.

Diffuse idiopathic skeletal hyperostosis (DISH) is a skeletal disease characterized by ossification of the anterior longitudinal ligament of the spine and various extraspinal ligaments. It causes stiffness and decreased range of spinal motion. The peripheral joints most commonly affected are the metacarpophalangeal joints, elbows, and shoulders. Diabetes, obesity, hypertension, and dyslipidemia are risk factors for this condition.

Hyperuricemia and gout are disorders associated with the metabolic syndrome. Thus, it is not surprising that type 2 diabetes patients are at increased risk for acute gout as well as chronic tophaceous gout.

Bursitis, particularly of the shoulders and hips, occurs more frequently than expected in patients with diabetes.

7. INFECTION

There are also several unusual infections that occur almost exclusively in diabetic patients (eg, emphysematous cholecystitis, mucormycosis, malignant otitis externa, and necrotizing

papillitis). As noted earlier, atherosclerosis with peripheral vascular disease is very common in the diabetic population, and the resultant ischemia undoubtedly plays a role in the frequent and severe lower extremity infections seen in these patients.

MANAGEMENT OF DIABETES IN THE HOSPITALIZED PATIENT

A number of studies have observed that hospitalized patients with diabetes and those with new-onset hyperglycemia (ie, those without a preadmission diagnosis of diabetes) have higher inpatient morbidity and mortality. The morbidity and mortality in diabetics is twice that of nondiabetics. Those with new-onset hyperglycemia have even a higher mortality—almost eightfold greater in one study. These observations have led to increased interest in improving glucose control in the hospitalized diabetic patient.

Most patients with diabetes are hospitalized for reasons other than their diabetes. Up to 10% to 15% of all hospitalized patients have diabetes. Audits suggest that as many as a 30% of hospitalized diabetic patients have inappropriate management of their diabetes such as being given metformin where contraindicated, failure to act on high blood glucose levels, omission of diabetes medication, no record of diabetes complications, inappropriate insulin management, or infrequent blood glucose monitoring. Use of outpatient oral therapies or insulin regimens in the hospital is challenging. Patients are not eating according to their normal schedule and are often fasting for procedures. It is, therefore, usual to use insulin, subcutaneous or intravenous, in the hospitalized patient. Insulin is safe to use in patients with cardiac, renal, and liver disease and its dosing can be adjusted to match changing inpatient needs.

Surgery represents a stressful situation during which most of the insulin antagonists (catecholamines, GH, corticosteroids) are mobilized. In the patient with diabetes, this can lead to a worsening of hyperglycemia and perhaps even ketoacidosis. The aim of medical management of people with diabetes during the perioperative period is to minimize these stress-induced changes. Recommendations for management depend both on the patient's usual diabetic regimen and on the type of surgery (major or minor) to be done.

For people with diabetes controlled with diet alone, no special precautions are necessary unless diabetic control is markedly disturbed by the procedure. If this occurs, small amounts of short-acting insulin, as needed, will correct the hyperglycemia. Patients on oral agents should not take them on the day of surgery. If there is significant hyperglycemia, short-acting insulin can be given as needed. If this approach does not provide adequate control, an insulin infusion should be started in the manner indicated later. The oral agents can be restarted once the patient is eating normally after the operation. It is important to order a postoperative serum creatinine level to ensure normal renal function prior to restarting metformin therapy.

Patients taking insulin represent the only serious challenge to management of diabetes when surgery is necessary. The insulin regimen used to control the glucose depends on the kind of diabetes (type 1 or type 2); whether it is minor surgery (<2 hours and patient eating afterward) or major surgery (longer than 2 hours, invasion of body cavity, not eating afterward); and the preoperative insulin regimen (eg, basal bolus or premixed insulin twice a day or premeal bolus only or regular before meals and NPH at bedtime). Type 1 patients have to be on some insulin to prevent the development of diabetic ketoacidosis. Many type 2 patients on insulin are fine without insulin for a few hours. Ideally patients with diabetes should undergo surgery early in the morning. Table 17–22 summarizes the approach for these various kinds of patients. One insulin infusion method adds 10 U of regular insulin to 1 L of 5% dextrose in half-normal saline, and this is infused intravenously at a rate of 100 to 180 mL/h. This gives the patient 1 to 1.8 U of insulin per hour, which, except in the most severe cases, generally keeps the blood glucose within the range of 100 to 250 mg/dL (5.5-13.9 mmol/L). The infusion may be continued for several days if necessary. Plasma glucose or blood glucose should be determined every 2 to 4 hours to be sure metabolic control is adequate. If it is not, adjustments in the ratio of insulin to dextrose in the intravenous solution can be made.

An alternative method, which is gaining increasing popularity, consists of separate infusions of insulin and glucose delivered by pumps to permit independent adjustments of each infusion rate, depending on hourly variation of blood glucose values. There are a number of different algorithms available for insulin infusions. Table 17–23 provides guidelines for management with an insulin infusion and an algorithm designed to achieve glycemic control in the range of 120 to 180 mg/dL blood glucose.

TABLE 17–22 Recommendations for management of insulin-treated diabetes during surgery.

	Minor Surgical Procedures (2 h; Eating Afterward)	Major Surgical Procedures (>2 h; Invasion of Body Cavity; Not Eating Immediately After Recovery)
Type 2 patients on basal bolus insulin regimen; twice daily premixed insulin	No insulin on the day of operation. Start 5% dextrose infusion; monitor fingerstix blood glucose and give subcutaneous short-acting insulin every 4 or 6 h	Same regimen as minor procedure. If control is not satisfactory, then intravenous insulin infusion
Type 1 patient on basal bolus insulin regimen or on insulin pump	Patients on pump should discontinue the pump the evening before procedure and should be given 24 h basal insulin. On day of procedure, start 5% dextrose; monitor blood glucose and give subcutaneous short-acting insulin every 4 or 6 h	Initiate insulin infusion on morning of procedure and transition back to usual regimen when eating

TABLE 17–23 Guideline for perioperative diabetes management with an intravenous insulin infusion.

1. **Maintenance of IV fluids (IV dextrose infusion must be maintained while the patient is on insulin infusion. Minimum rate of 10 mL/h)**
   ☐ D5 NS at 100 mL/h
   ☐ D5 1/2 NS at 100 mL/h
   ☐ D10 NS at _____ mL/h (for patients with fluid restrictions or renal failure)
   ☐ Additive: KCl 20 mEq/L (*generally 20 mEq/L*)
   ☐ Other_____ at _____ mL/h
2. **Regular insulin infusion 100 U regular insulin in 100 mL NS (1 U = 1 mL)**
   A. Flush first 20 mL of infusion through tubing before connecting to patient
   B. Before beginning infusion, check blood glucose (BG) with glucose meter
3. **Start insulin infusion rate as follows (when BG ≥100 mg/dL)**
   ☐ 0.3 U/h taking <30 U insulin daily (*recommended for type 1; Pancreatectomy*)
   ☐ 1 U/h for patients previously diet controlled, taking oral hypoglycemic agent, or <30 U insulin daily
   ☐ 1.5 U/h for patients taking >30 U insulin daily
   ☐ Other _____ U/h
4. **Adjust insulin infusion rate as follows:**

   ☐ **Standard adjustment**   ☐ **Sensitive Adjustment (for Type 1; Pancreatectomy)**
   BG <80 mg/dL; stop infusion and **call MD;**   BG <80 mg/dL; stop infusion and **call MD;**
   see instruction #5 below   **see instruction #5 below**
   Do not restart insulin infusion   Do not restart insulin infusion
   until BG >100 mg/dL[a]   until BG >100 mg/dL[a]
   BG 80-120; decrease drip by 0.5 U/h   BG 80-120; decrease drip by 0.2 U/h
   BG 121-180; no change in drip rate   BG 121-180; no change in drip rate
   BG 181-250; increase drip by 0.5 U/h   BG 181-250; increase drip by 0.2 U/h
   BG >250 bolus 5 U regular insulin IV and   BG >250 bolus 2 U regular insulin IV and
   increase drip by 0.5 U/h   increase drip by 0.2 U/h

5. **For a BG <80 mg/dL or >400 mg/dL on insulin infusion, call MD**
   ☒ BG <80 mg/dL but >60 mg/dL, stop insulin infusion; check BG every 15 min
   ☒ BG ≤60 mg/dL, stop insulin infusion; give **50-mL D50W IV push**; check BG every 15 min and repeat treatment until BG ≥100 mg/dL. When BG ≥100 mg/dL, call MD for new insulin infusion rate
   ☒ BG >400 mg/dL, call MD to reassess insulin infusion rate
   ☒ If TPN or tube feeds are interrupted for longer than 30 min, start D10W at 50 mL/h. Notify MD about change and future action
6. **When converting to subcutaneous (SQ) insulin, give prescribed SQ dose 30 min prior to discontinuing insulin infusion. Use Adult** SQ Insulin Order Sheet
7. **If patient eating meals give _____ U aspart SQ after patient eats carbohydrates and continue insulin infusion**
8. **Discontinue insulin infusion maintenance IV fluids when insulin infusion discontinued**

[a]Check BG every hour with glucose meter until stable (range 100-180 mg/dL) for two consecutive readings and then every 2 hours.

After surgery, when the patient has resumed an adequate oral intake, intravenous administration of insulin and dextrose can be stopped half an hour after the first subcutaneous insulin injection. Insulin needs may vary in the first several days after surgery because of continuing postoperative stresses and because of variable caloric intake. In this situation, multiple doses of regular insulin, guided by blood glucose determinations, can keep the patient in acceptable metabolic control.

In the intensive care units, glucose levels are controlled most frequently using insulin infusions. Patients on total parental nutrition (TPN) can have insulin added to the bag. Standard TPN contains 25% dextrose, so 50-mL/h infusion delivers 12.5 g of dextrose per hour.

On the general surgical and medical wards, most patients are managed on subcutaneous insulin regimens. Limited cross-sectional and prospective studies suggest that the best glucose control is achieved on a combination of basal and bolus regimen with 50% of daily insulin needs provided by intermediate or long-acting insulins. Standardized order sets prompt medical personnel to write more physiological insulin orders, reduce errors, and include algorithms for recognition and treatment of hypoglycemia (see http://ucsfinpatientdiabetes.pbworks.com for examples). Table 17–24 is an example of one such order set for a subcutaneous insulin regimen.

Targets for Glucose Control in the Hospitalized Patient

A number of studies have observed that hospitalized patients with diabetes and those with new-onset hyperglycemia (ie, those without a preadmission diagnosis of diabetes) have higher inpatient morbidity and mortality. Also, poor glucose control in diabetics at admission is associated with increased size of myocardial infarction and in some studies increased nosocomial infections after surgery. These observations have led to the question of whether lowering glucose levels close to normal in the hospital improves outcomes.

Initial studies of patients in surgical and medical ICUs suggested that aggressive treatment of hyperglycemia (blood glucose <110 mg/dL) reduced mortality and morbidity. A large multicenter, multinational study (NICE-SUGAR), however, failed to confirm these reports. This study recruited 6104 surgical and medical ICU patients with hyperglycemia (20% had diabetes) and randomized them to tight control (81-108 mg/dL) or less tight control (<180 mg/dL). The tight group achieved blood glucose levels of 115 ± 18 and the conventional group 144 ± 23. There were more deaths (829 vs 751) in the tight glucose control group compared to the less tight glucose control group ($p = 0.02$). The intensively treated group also had more cases of severe

TABLE 17–24 Example of a standardized subcutaneous insulin order set for inpatient use.

1. Check blood glucose and give insulin before meals, bedtime, and 2 AM.

2. Discontinue previous SQ insulin order.

3. If patient becomes NPO for procedure/stops eating:

• **HOLD** nutritional dose of rapid-acting analog

• Give correctional dose of rapid-acting analog if BG >130 mg/dL

• Give glargine dose. If BG has been <70 mg/dL in last 24 h, call MD to consider adjusting glargine dose

• Call MD for NPO orders if patient on 70/30, NPH insulin or has been NPO >12 h

A. Basal and nutritional insulin dose (in units):

Time	Breakfast	Lunch	Dinner	Bedtime
Rapid-acting analog				
NPH				
Insulin glargine				
Insulin mixture				

B. Meal time correctional insulin with rapid-acting analog. Check box to choose scale. Add or subtract from nutritional dose of rapid-acting analog.

Blood Glucose Range	☐ Sensitive BMI <25 and/or <50 U/d	☐ Average BMI 25-30 and/or 50-90 U/d	☐ Resistant BMI >30 and/or >90 U/d	☐ Custom
<70 mg/dL Once BG • 100, give	Treat for hypoglycemia per protocol (see order #6). Once BG •100 mg/dL give rapid-acting analog with following change when patient eats:			
	2 U less	3 U less	3 U less	- - - U less
70-100	1 U less	2 U less	3 U less	- - - U less
101-130	Give nutritional rapid-acting insulin as in #4A			
131-150	+ 0	+ 1	+ 2	+ - - -
151-200	+ 1	+ 2	+ 3	+ - - -
201-250	+ 2	+ 4	+ 6	+ - - -
251-300	+ 3	+ 6	+ 9	+ - - -
301-350	+ 4	+ 8	+ 12	+ - - -
351-400	+ 5	+ 10	+ 15	+ - - -
>400	+ 6	+ 12	+ 18	+ - - -

C. ☒ Bedtime and 2 AM correctional insulin with rapid-acting analog if BG ≥200 mg/dL.

BG Range (mg/dL)	Default Value (U)	Custom
200-250	1	- - - U
251-300	2	- - -
>300	3	- - -

4. Call MD for BG <70 mg/dL or >400 mg/dL.

5. For BG <70 mg/dL, use Hypoglycemia Protocol below. These hypoglycemia orders remain active for duration of SQ insulin administration. For patient taking PO, give 20 g of oral fast-acting carbohydrate per patient preference:

☒ Give 4 glucose tablets (5 g glucose/tablet). Repeat Q 15 min until BG ≥100 mg/dL.

or

☒ Give 6 oz fruit juice. Repeat Q 15 min until BG ≥100 mg/dL.

☒ Give 25-mL D50W IV push if patient cannot take PO. Repeat Q 15 min until BG ≥100 mg/dL.

☒ Check fingerstick glucose every 15 min and repeat treatment until BG is ≥100 mg/dL.

6. Discontinue above monitoring and intervention orders when SQ insulin is discontinued.

Note: Glargine (Lantus) *cannot* be mixed with any other insulin. Give glargine as a separate injection.

hypoglycemia (206 vs 15 cases), and the excess deaths in the intensive group were due to cardiovascular events.

A study on tight intraoperative glycemic control during cardiac surgery also failed to show any benefit; if anything, the intensively treated group had more events. The United Kingdom Glucose Insulin in Stroke Trial (GIST-UK) failed to show beneficial effect of tight glycemic control in stroke patients; however, the investigators acknowledged that, because of slow recruitment, the study was underpowered.

Thus, based on the evidence available so far, ICU patients with diabetes and new-onset hyperglycemica with blood glucose levels above 180 mg/dL should be treated aiming for target glucose levels between 140 mg/dL (7.8 mmol/L) and 180 mg/dL (10 mmol/L). Targeting blood glucose control in the intensive care unit close to 100 mg/dL (5.6 mmol/L) has not been proven to be beneficial and may even be harmful. When patients leave the ICU, target glucose values between 100 mg/dL and 180 mg/dL may be appropriate, although this view is based on clinical observations rather than conclusive evidence.

DIABETES MELLITUS AND PREGNANCY

Hormone and Fuel Balance During Pregnancy

During pregnancy, the rapidly growing fetus impacts metabolism in the mother and causes profound hormonal and metabolic changes. These changes significantly affect the management of preexisting diabetes, and can precipitate hyperglycemia and diabetes in previously nondiabetic mothers.

Early in pregnancy, the hormones of pregnancy initiate changes that prepare the mother for the increasing nutrient requirements of the fetus late in pregnancy. Fat deposition is accentuated in early pregnancy due to enhanced conversion of glucose to triglyceride after meals in pregnant compared with nonpregnant subjects. Elevated prolactin and placental lactogen drive the production of serotonin by the β cell, which leads to β cell expansion, a process that peaks in mid gestation. β cell mass approximately doubles during pregnancy, and both β cell glucose sensitivity and insulin secretory capacity increase. At the same time, α cell mass and glucagon secretion remain unchanged.

As the pregnancy progresses and nutrient utilization by the fetus increases, plasma concentrations of glucose in the fasting state tend to decline slightly. Since the flow of glucose and other nutrients to the fetus depends on the gradient across the placenta, maintenance of maternal glucose levels becomes increasingly important and is achieved through increasing insulin resistance in the liver, muscle, and adipose tissue. Several pregnancy hormones contribute to insulin resistance, but progesterone, which increases steadily through pregnancy and peaks shortly prior to parturition, provides the major drive. Glucose availability for the fetus increases as increased lipolysis in adipose tissue leads to a shift toward fatty acid utilization by maternal tissues. Ketogenesis is also accentuated in the postabsorptive state during pregnancy, secondary to hormonal effects on the maternal liver cells and

increased provision of substrate FFA. The increased FFAs contribute to the marked reduction in insulin-mediated glucose uptake by skeletal muscle characteristic of late pregnancy. As a net result, during the third trimester, pregnant women without diabetes run slightly higher blood glucose and ketone levels than nondiabetic subjects, and women with preexisting diabetes that do not adjust therapy appropriately can develop severe hyperglycemia.

The fetus also plays a role in the maintenance of the maternal to fetal glucose gradient—the fetal glucose levels being determined by fetal insulin levels and the fetal insulin sensitivity. Maternal hyperglycemia due to inadequate control of maternal diabetes early in pregnancy will lead to an increase in fetal glucose levels, fetal hyperinsulinemia, and large for gestational age fetus.

Pregnancy in Women with Preexisting Diabetes

A. Prevalence The prevalence of pregestational diabetes in pregnant women is increasing. In the United Kingdom, the prevalence of type 1 diabetes increased from 1.56 to 4.09 per 1000 pregnancies between 1995 and 2015. For type 2 diabetes, the prevalence increased from 2.34 to 5.09 per 1000 pregnancies between 1995 and 2008 with a more rapid increase to 10.62 per 1000 pregnancies by 2012.

B. Maternal, fetal, and neonatal consequences of presence of diabetes during pregnancy If diabetes is poorly controlled in the first weeks of pregnancy, the risks of spontaneous abortion and congenital malformation of the infant are increased. Poor glycemic control later in pregnancy is associated with still births, fetal macrosomia, polyhydramnios, and neonatal hypoglycemia. Preexisting diabetes-related complications, particularly gastroenteropathy, retinopathy, and nephropathy can impact both the pregnant woman and the fetus.

Congenital anomalies Poorly controlled diabetes in the first weeks of pregnancy raises the risks of spontaneous abortion and congenital anomalies of the infant. The commonest anomalies are cardiac, neural tube defects, and genitourinary abnormalities. Their incidence compared to the normal population and their presumed time of occurrence during embryonic development are listed in Table 17–25. The rate of anomalies is higher in diabetic women with poor preconception control. Observational studies demonstrate a linear relationship between glycated hemoglobin levels and malformation rates. From a systematic review of several observational studies, it was estimated that the relative risk reduction of congenital anomalies for each 1% decrease in HbA_{1c} ranged from 0.39 to 0.59. Attending preconception clinics with intensive diabetic management instituted prior to conception and continued through early pregnancy, results in significant reduction in the frequency of anomalies—to rates near the population norms.

Polyhdramnios Later in pregnancy, polyhydramnios, a common complication in women with poorly controlled diabetes, may lead to preterm delivery. Polyhydramnios also increases the risk for

TABLE 17–25 Congenital malformations in infants of diabetic mothers.[a]

	Ratio of Incidences Diabetic vs Control Group	Latest Gestational Age for Occurrence (Weeks after Menstruation)
Caudal regression syndrome (due to premature termination of the vertebral column—resulting in agenesis of the sacrum and coccyx and/or malformations of the pelvis and legs)	252	5
Anencephaly	3	6
Spina bifida, hydrocephalus, or other central nervous system defects	2	6
Cardiac anomalies	4	
Transposition of great vessels		7
Ventricular septal defect		8
Atrial septal defect		8
Anal/rectal atresia	3	8
Renal anomalies	5	
Renal agenesis	6	7
Cystic kidney	4	7
Ureter duplex	23	7
Situs inversus	84	6

[a]Data from Kucera J. Rate and type of congenital anomalies among offspring of diabetic women. *J Reprod Med*. 1971;7:61; and Mills JL, Baker L, Goldman AS. Malformations in infants of diabetic mothers occur before the seventh gestational week: implications for treatment. *Diabetes*. 1979;28:292.

placenta abruptio and postpartum uterine atony. Polyhydramnios is an excess volume of amniotic fluid (>1000 mL, often >3000 mL). It is most often associated with fetal macrosomia. The excess volume of amniotic fluid is not related simply to the concentration of glucose or other solutes in amniotic fluid or to excess fetal urine output as measured by change in bladder size on ultrasonography. Other possible factors include decreased fetal swallowing, decidual and amniotic fluid prolactin, and as yet unknown determinants of the complicated multicompartmental intrauterine transfer of water. Polyhydramnios is rare in women with well-controlled diabetes.

Fetal macrosomia   Many fetuses of poorly controlled diabetic mothers are **macrosomic** (birth weight >90th percentile for gestational age), with increased fat stores, increased length, and increased abdomen-to-head or thorax-to-head ratios. The hypothesis that fetal macrosomia results from the causal chain of maternal hyperglycemia → fetal hyperglycemia → fetal hyperinsulinemia → fetal macrosomia has been confirmed by clinical and experimental studies. Macrosomic infants of diabetic mothers have significantly higher concentrations of C peptide (representing endogenous insulin secretion) in their cord sera or amniotic fluid than in those with birth weights appropriate for gestational age. Monkey fetuses with insulin-releasing pellets implanted in utero become macrosomic. Other metabolic substrates that cross the placenta, such as branched-chain amino acids may also play a role in fetal macrosomia, and transplacental lipids could contribute to fat deposition.

Prevention of maternal hyperglycemia throughout pregnancy can reduce the incidence of macrosomia. The glycemic threshold for fetal macrosomia seems to be *postprandial* peak values above 130 mg/dL (7.2 mmol/L). On the other hand, excessively tight glycemic control (average peak postprandial blood glucose levels <110 mg/dL [6.1 mmol/L]) can be associated with insufficient fetal growth and small-for-date infants.

Complications of macrosomia include fetopelvic disproportion leading to shoulder dystocia and its attendant risk for brachial plexus injury and humeral and clavicle fractures. The neonate with macrosomia is also at increased risk for hypoglycemia, hyperbilirubinemia, hypocalcemia, respiratory distress syndrome, and polycythemia.

Adverse maternal outcomes of difficult vaginal delivery include severe perineal lacerations and subsequent urinary and/or fecal incontinence.

Intrauterine growth retardation   The fetus of a woman with diabetes of long duration and vascular disease may suffer intrauterine fetal growth restriction related to inadequate uteroplacental perfusion. All body diameters may be below normal on ultrasonographic measurements, but the abdominal circumference is especially affected, and oligohydramnios and abnormal Doppler flow measurements of the umbilical cord are common. In these patients provision of adequate rest, meticulous control of hypertension (target <135/85 mm Hg), maintenance of normal blood glucose levels, and intensive fetal surveillance are all essential for success.

Intrauterine death   Prior to the 1970s, the incidence of apparently sudden intrauterine fetal demise in the third trimester of diabetic pregnancies was at least 5%. Except for congenital malformations, the cause of stillbirth is often not obvious. The risk is greater with poor diabetic control, and the incidence of fetal death exceeds 50%, if ketoacidosis develops in the mother. Some instances of fetal demise are associated with preeclampsia-eclampsia, which is more common in pregnancies complicated by diabetes. Fetal death has also been associated with pyelonephritis—this is now largely prevented by screening for and treating asymptomatic bacteriuria. Other than these known risk factors, one can presume (based on experimental studies) that the combination of fetal hyperglycemia and hypoxia leads to acidosis and myocardial dysfunction. Good glycemic control in diabetic women greatly reduces the risk of stillbirth.

Neonatal hypoglycemia   Hypoglycemia is common in the first 48 hours after delivery from previously hyperglycemic mothers and is defined as blood glucose below 36 mg/dL (2.0 mmol/L) regardless of gestational age. The risk is as high for infants of mothers with type 2 diabetes as type 1 diabetes. Neonatal

hypoglycemia is most closely linked to poor maternal glucose control and elevated fetal insulin levels during labor and delivery, although data from amniotic fluid insulin studies suggest that the more hyperinsulinemic the fetus in the late third trimester, the higher the risk of neonatal hypoglycemia. Infants of diabetic mothers may also have deficient catecholamine and glucagon secretion, and the hypoglycemia may be related to diminished hepatic glucose production and oxidation of free fatty acids. The symptomatic infant may be lethargic with associated apnea, tachypnea, cyanosis, or seizures. Early feeding with 10% dextrose in water by bottle or gavage by 1 hour of age should be instituted in the at-risk neonate. If this is not successful, treatment with intravenous dextrose solutions is indicated. There are usually no long-term sequelae of episodes of neonatal hypoglycemia. Keeping maternal blood glucose levels below 144 mg/dL and tight glycemic control during labor can reduce the frequency of neonatal hypoglycemia.

Diabetic gastroenteropathy   In early gestation, diabetic gastroparesis can severely exacerbate the nausea and vomiting of pregnancy (hyperemesis gravidarum), which sometimes continues into the third trimester. Drugs stimulating gastric motility such as erythromycin may be useful, but many patients with this complication require hyperalimentation to achieve nutritional intake adequate for fetal development.

Diabetic retinopathy   Background diabetic retinopathy may develop or progress during pregnancy, but it usually regresses postpartum. If background retinopathy is already present in early pregnancy, the rate of progression to neovascularization over the course of the pregnancy (proliferative diabetic retinopathy) is 6% if the background retinopathy is mild; 18% if it is moderate; and 38% if severe preproliferative changes are present. The risk factors for progression to proliferative retinopathy include poor glycemic control before and during early pregnancy, rapid improvement in glycemic control during pregnancy, hypertension, and perhaps the many growth factors derived from placental tissue. These risks are an important reason to institute intensified preconception management of diabetes. During pregnancy, sequential ophthalmologic examinations are essential in women with type 1 or type 2 diabetes, and laser photocoagulation treatment of the retina may be necessary.

Diabetic nephropathy   The risk of worsening of diabetic nephropathy during pregnancy depends on baseline renal function and the degree of hypertension. Total urinary albumin excretion does not increase substantially in normal pregnancy, but total urinary protein collections, which obstetricians have used to define preeclampsia, may show a twofold increase in uncomplicated gestation. Diabetic women with microalbuminuria (30-299 mg/24 h) may have worsening of the albuminuria during pregnancy and 15% to 45% develop the preeclamptic syndrome. Albuminuria usually regresses postpartum. Women with macroalbuminuria (24-hour urinary albumin >300 mg) and well-preserved renal function (serum creatinine <1.2 mg/dL [<106 µmol/L]; creatinine clearance >80 mL/min) at the beginning of pregnancy are likely to have moderate decline in renal function during gestation, and

approximately 6% may have renal failure at follow-up several years after pregnancy. The latter figure may not be different from the course of diabetic nephropathy in nonpregnant women with this level of renal dysfunction. If initial renal function in pregnancy is impaired (serum creatinine >1.2 mg/dL [>106 µmol/L]; creatinine clearance <80 mL/min), then 35% to 40% are expected to show further decline during pregnancy, and 45% to 50% have renal failure at follow-up several years later. Thus, careful preconception counseling is important for these patients and their family members.

Management

Pregnant women with preexisiting diabetes should ideally receive care in a multidisciplinary clinic staffed by obstetricians, endocrinologists, diabetes nurse specialists, dietitians, and specialist midwives. There is substantial evidence that this improves the outcomes in women with pregestational diabetes.

Preconception counseling   All diabetic women of reproductive age should be counseled to use adequate birth control and plan their pregnancies. Before any planned pregnancy, the couple should be referred for preconception assessment and counseling. There should be assessment and discussion regarding management of diabetic complications. Baseline measurements should include thyroid function tests including autoantibodies; early morning spot urine albumin/creatinine ratio; urinalysis to rule out asymptomatic bacteriuria; retinal examination; and cardiological assessment in high-risk patients such as those with evidence of microvascular disease. ACE inhibitors and ARBs cannot be used during pregnancy. If necessary, alternative antihypertensive agents that are safe for use in pregnancy (methyldopa, nifedipine, amlodipine, labetolol) should be prescribed. Lipid-lowering therapies are also contraindicated, and lipid abnormalities must be managed by dietary measures. There should be discussion of risks and expected management strategies in the pregnancy. The patient should be given glycemic goals with the aim of achieving control as close to normal as possible before conception.

Glucose and insulin management   Women are seen every 2 to 4 weeks in the first and second trimester and then every 1 to 2 weeks until 36 weeks and then weekly to term. The goal of glucose management during pregnancy is to prevent both preprandial and postprandial hyperglycemia. Patients with type 2 diabetes on oral agents are usually switched to insulin treatment during pregnancy. There are small nonrandomized studies suggesting that metformin use is safe in the first trimester. The drug does cross the placenta, but it has not been associated with teratogenesis. Glyburide was thought not to cross the placenta and has been considered safe for use in pregnancy. More recently it has been found that glyburide does cross the placenta and fetal glyburide levels are approximately 70% of maternal plasma levels. Until larger randomized control studies of metformin and glyburide use at conception and early pregnancy are available, caution should be exercised regarding their use early in pregnancy.

Perinatal outcome is optimal if patients aim for fasting plasma glucose levels below 100 mg/dL (5.6 mmol/L) and

postprandial levels below 130 mg/dL (7.2 mmol/L). For patients with a history of recurrent severe hypoglycemic reactions, somewhat higher blood glucose targets should be selected. Self-monitoring of capillary blood glucose should be performed eight or more times a day. Occasional monitoring in the middle of the night is recommended to monitor for nocturnal hypoglycemia. There is evidence that CGM systems may be helpful in reducing both hyperglycemia and hypoglycemic excursions. These systems do not replace self-monitoring of blood glucose but allow for fine-tuning of glycemic control and alert the patient to rapid changes in glucose levels. One randomized study showed that pregnancies using CGM had better glycemic control in the third trimester and lower risk of macrosomia. Clinical trials using continuous glucose monitors during pregnancy are currently underway. Confirmation of long-term control is provided by sequential measurement of glycosylated hemoglobin and fructosamine every 4 to 6 weeks.

Patients should see the nutritionist to assess caloric needs and get instructed on carbohydrate counting. The caloric intake is based on ideal body weight and it is approximately 30 to 35 kcal/kg for normal-weight women; about 24 kcal/kg current weight for overweight women; and 10 to 15 kcal/kg current weight for obese women. All patients should learn how to self-adjust their doses of short-acting insulin based on planned carbohydrate load or pre-meal blood glucose levels.

Type 1 patients are typically on a basal-bolus insulin regimen (see Table 17–15). Regular insulin, insulin aspart, and insulin lispro are the approved bolus insulins; and NPH and insulin detemir are used for basal coverage. Studies on insulin glargine use during pregnancy are limited. Retrospective case control studies have not shown any adverse effects of glargine use at the time of conception and during pregnancy. Many type 1 patients elect to use insulin pumps during pregnancy. Since the risk of diabetic ketoacidosis is increased with pump use, it is especially critical that patients have adequate instruction on the use of insulin pumps and can troubleshoot any problems that may arise.

Women with type 2 diabetes can be managed with regular insulin or short-acting insulin analogs before meals and NPH or insulin detemir at bedtime. Sometimes NPH or insulin detemir is also required in the morning.

Total insulin requirements vary during gestation. There is usually an increase in insulin dose between weeks 3 and 7, then a slight decrease between weeks 7 and 15 followed by a gradual increase until about week 35. The insulin requirements in type 1 diabetes at around 35 weeks (1.0 U/kg/24 h) are almost double the prepregnancy requirements. Women with type 2 diabetes usually start out with higher doses and may eventually require as much as 1.5 to 2 U/kg/24 h. Exercise can improve insulin sensitivity and should be encouraged in type 2 patients. The benefits of exercise are less obvious in women with type 1 diabetes, and there is always a concern about exercise-induced hypoglycemia.

Hypoglycemic reactions are more frequent and sometimes more severe in early gestation but are a risk at any time during pregnancy. Therefore, insulin-treated patients should take snacks between meals and at bedtime to prevent hypoglycemia. Family members should be instructed in the use of glucagon. The use of

TABLE 17–26 Schedule of obstetric tests and procedures.

Procedure	Risk Based on Glycemic Control, Presence of Vascular Disease	
	Low Risk	**High Risk**
Ultrasound to date gestation	8-12 wk	8-12 wk
Prenatal genetic diagnosis	As needed	As needed
Targeted perinatal ultrasound; fetal echocardiography	18-22 wk	18-22 wk
Fetal kick counts	28 wk	28 wk
Ultrasound for fetal growth	28 and 37 wk[a]	Every 3-8 wk
Antepartum FHR monitoring, backup with biophysical profile	36 wk, weekly	27 wk, 1-3 per wk
Amniocentesis for lung	…	35-38 wk
Induction of labor	41 wk[b]	35-38 wk

[a]Not needed in normoglycemic, diet-treated women with gestational diabetes mellitus.
[b]Earlier for obstetric reasons or for impending fetal macrosomia.

continuous glucose monitors is encouraged. Hypoglycemic reactions have not been associated with fetal death or congenital anomalies, but they pose a risk to maternal health.

Fetal monitoring (Table 17–26)   First-trimester ultrasound is often obtained to document viability, as the rate of miscarriage is higher in women with diabetes, and to estimate gestational age. A second-trimester, detailed ultrasound examination of fetal anatomy is performed at approximately 18 weeks for detection of congenital anomalies. Ultrasound examination in the third trimester (at 28-32 weeks and 38 weeks) is performed to assess fetal growth.

Antepartum fetal testing is recommended using fetal movement counting, biophysical profile, nonstress test (NST), and/or contraction stress test (CST) initiated at 32 to 34 weeks of gestation. Most simply, the infrequency of fetal movement as noted in regular fetal kick counts (fewer than four per hour) may indicate fetal jeopardy. The fetal biophysical profile is a rigorous analysis of fetal activity patterns using ultrasound to assess gross body movements, the tone of the limbs, and chest wall motions as well as reactivity of the fetal heart rate (FHR) and the volume of amniotic fluid. The presence of FHR accelerations and long-range variability on the NST and the absence of late decelerations (lower FHR persisting after the contraction subsides) on the CST indicate that the fetus is well oxygenated. Generally, the NST and CST are sensitive screening tests, and abnormal results of FHR monitoring in these tests overestimate the diagnosis of fetal distress. Therefore, it is wise to obtain additional evidence of fetal jeopardy (by biophysical ultrasonographic assessment) before cesarean delivery is recommended in preterm pregnancies. In term gestation with abnormal fetal testing, there is little to be gained by continuing the pregnancy. In complicated patients with IUGR, oligohydramnios, preeclampsia, or poor blood glucose control, testing may start as early as 26 weeks and is performed more frequently.

Timing of delivery   Unless maternal or fetal complications arise, the goal for delivery in diabetic women should be 38 to 41 weeks in order to reduce neonatal morbidity from preterm deliveries. On the other hand, the obstetrician may wish to induce labor before 39 weeks, if there is concern about increasing fetal weight. Before a preterm delivery decision (<37 weeks) is made or delivery is considered in women with poor glycemic control at 37 to 38 weeks of gestation, fetal pulmonary maturity should be determined. Tests for maturity using amniocentesis assess risk of neonatal respiratory distress syndrome and include the lecithin-sphingomyelin (L/S) ratio, phosphatidylglycerol, and other biochemical or physical assays of surfactant activity. In pregnancies complicated by hyperglycemia, fetal hyperinsulinemia can lead to low pulmonary surfactant apoprotein production. The lowest risk of respiratory distress syndrome is attained by delaying delivery (if possible) until 38 to 41 weeks and minimizing the need for cesarean sections.

Route of delivery   Once fetal lung maturity is likely, the route of delivery must be selected based on the usual obstetric indications. If the fetus seems large (>4200 g) on clinical and ultrasonographic examination of diabetic women, cesarean section probably should be performed because of the possibility of shoulder dystocia and birth trauma. Otherwise, induction of labor is reasonable, because maternal and peripartum risks are fewer following vaginal delivery. Once labor is underway, continuous FHR monitoring is essential. Maternal blood glucose levels greater than 150 mg/dL (8.3 mmol/L) can be associated with intrapartum fetal hypoxia.

Insulin and glucose management during labor and delivery   It is important to avoid maternal hyperglycemia during labor because it increases the risk for fetal acidemia and neonatal hypoglycemia. The goal is to maintain glucose levels between 70 and 110 mg/dL (3.9-6.1 mmol/L). Maternal glucose levels above 180 mg/dL (10 mmol/L) are consistently associated with neonatal hypoglycemia. During the latent phase of labor, maternal metabolic demands are fairly stable but there is significant energy expenditure during active labor and delivery and insulin needs are minimal. Glucose control is typically managed using intravenous dextrose (at 100-125 mL/h) and insulin infusions (Table 17–27).

After delivery of the placenta, the insulin-resistant state rapidly disappears, and insulin requirements are close to prepregnancy levels. Type 1 or type 2 patients who were previously on insulin can go back to their usual prepregnancy insulin regimens and doses once they start eating. Type 2 patients who were on oral agents prepregnancy frequently do not require any medication during the first 24 to 48 hours postpartum. They can stay on insulin while breastfeeding or go on metformin or glyburide which are safe while breastfeeding.

Neonatal management   Planning for the care of the infant should be started prior to delivery, with participation by the pediatrician or neonatologist in decisions about timing and management of delivery. In complicated cases, the pediatrician must be in attendance to learn about antenatal problems, to assess the need for resuscitation, to identify major congenital anomalies, and to

TABLE 17–27   Protocol for intrapartum insulin infusion.[a]

Intravenous Fluids
If blood glucose is >130 mg/dL (>7.2 mmol/L), infuse mainline Ringer lactate at a rate of 125 mL/h.
If blood glucose is <130 mg/dL (<7.2 mmol/L), infuse mainline Ringer lactate to keep vein open and begin Ringer lactate and 5% dextrose at a rate of 125 mL/h controlled by infusion pump.

Insulin Infusion
Mix 100 U of regular human insulin (U100) in 100 mL NaCl 0.9% and piggyback to mainline. The concentration is 1 U/mL. Adjust intravenous insulin hourly according to the following table when the blood glucose is >70 mg/dL (>3.9 mmol/L). Target blood glucose 70-110 mg/dL.

Blood Glucose in mg/dL (mmol/L)	Type 1	Type 2	Gestational Diabetes	Custom
<70 (<3.9)	No insulin	No insulin	No insulin	
71-90 (3.9-5)	0.5 U/h	No insulin	No insulin	
91-110 (5.1-6.1)	1	1 U/h	No insulin	
111-130 (6.2-7.2)	1.5	2	1 U/h	
131-150 (7.3-8.3)	2	3	2	
151-170 (8.4-9.4)	2.5	4	3	
171-190 (9.5-10.6)	3	5	4	
>190 (>10.6)	Call MD and check urine ketones			

[a]Protocol useful also for diabetic pregnant women who are "NPO" or being treated with beta-adrenergic tocolysis or corticosteroids. The scale dosages may need to be doubled for the latter. Boluses of short-acting insulin must be used to cover meals.

plan initial therapy for the sick infant if required. Infants of mothers with poorly controlled diabetes have an increased risk of respiratory distress syndrome. Possible reasons include abnormal production of pulmonary surfactant or connective tissue changes leading to decreased pulmonary compliance. However, in recent years, the incidence of respiratory distress syndrome in these pregnancies has declined from 24% to 5%, probably related to better maternal glycemic control, selected use of amniotic fluid tests, and delivery of most infants at term. Other possible problems in infants of diabetic mothers include hypocalcemia less than 7 mg/dL (1.75 mmol/L), hyperbilirubinemia greater than 15 mg/dL (256 μmol/L), polycythemia (central hematocrit >70%), and poor feeding. These complications are presumably related to fetal hyperglycemia and hyperinsulinemia and probably to intermittent low-level fetal hypoxia. Improved control of the maternal diabetic state has reduced their incidence.

Gestational Diabetes

Gestational diabetes (GDM) is defined as glucose intolerance that develops or is first recognized during pregnancy. As insulin resistance increases during pregnancy, euglycemia depends on a compensatory increase in insulin secretion. Failure to compensate with increased insulin secretion leads to gestational diabetes.

As the increase in insulin resistance is greatest in the third trimester, GDM usually develops going into this period. The prevalence of GDM in a population is reflective of the prevalence of type 2 diabetes in that population. In low-risk populations, such as those found in Sweden, the prevalence in population-based studies is lower than 2% even when universal testing is offered, while studies in high-risk populations, such as the Native American Cree, Northern Californian Hispanics, and Northern Californian Asians, reported prevalence rates ranging from 4.9% to 12.8%. Other risk factors include a history of macrosomia (birth weight >4000 g), polycystic ovarian syndrome, essential hypertension or pregnancy-related hypertension, history of spontaneous abortions and unexplained still births, family history of diabetes, obesity, age older than 25 years, and history of gestational diabetes (Table 17–28).

Women with GDM are at increased risk for urinary tract infections, pyelonephritis, asymptomatic bacteriuria, and preeclampsia. There is a 10% risk of polyhydramnios. Fetal risks of poor glucose control include stillbirths and macrosomia. There is no increase in risk for congenital anomalies since the glucose intolerance develops later in pregnancy.

Strategies for diagnosis are outlined in Table 17–29. Currently, international consensus is lacking regarding the diagnostic criteria for GDM. The 100 g 3-hour OGTT test is commonly used in the United States and the 75 g 2-hour WHO test is used in many other countries. The risk assessment for GDM (see Table 17–28) should be performed at the first prenatal visit. Those women at high risk should undergo OGTT testing as soon as feasible. If the test is negative, then they should be retested at 24 to 28 weeks of gestation. A fasting plasma glucose of more than or equal to 126 mg/dL or random glucose of more than or equal to 200 mg/dL confirmed on repeat testing is diagnostic of diabetes and negates the need to perform a glucose challenge test. Lower risk women are screened for GDM using a two-step protocol performed at 24 to 28 weeks. First, a glucose challenge test is performed using a 50 g glucose load. The patient does not have to be fasting. A 1-hour

TABLE 17–28   Risk factors for gestational diabetes mellitus.

Maternal age >37 y
Ethnicity
Indian subcontinent—11-fold
Southeast Asian—eightfold
Arab/Mediterranean—sixfold
Afro-Caribbean—threefold
Prepregnancy weight >80 kg or BMI >28 kg/m^2
Family history of diabetes in first-degree relative
Previous macrosomia/polyhydramnios
Previous unexplained stillbirth
Polycystic ovarian syndrome

TABLE 17–29   Screening and diagnosis of gestational diabetes.

Perform blood glucose testing at 24-28 wk

Two-step protocol:

A. **Glucose challenge test:** 1-h plasma glucose after 50 g glucose load. Patient is not fasting. ≥140 mg/dL (≥7.8 mmol/L) identifies ~80% of women with GDM; ≥130 mg/dL (≥7.2 mmol/L) increases sensitivity to ~90%.

B. **100 g OGTT:** If a woman meets or exceeds threshold on glucose challenge test then perform fasting diagnostic 100 g OGTT within a week. The test is performed after an at least 8-h overnight fast and after at least 3 d of unrestricted diet (≥150 g carbohydrate per day). Subject should be seated and should not smoke during test.

The test is positive if two of the following glucose levels are found.

	Carpenter/Coustan criteria	National Diabetes Data group criteria
Fasting:	≥95 mg/dL (≥5.3 mmol/L)	>105 mg/dL (5.8 mmol/L)
1 h:	≥180 mg/dL (≥10 mmol/L)	>190 mg/dL (10.6 mmol/L)
2 h:	≥155 mg/dL (≥8.6 mmol/L)	>165 mg/dL (9.2 mmol/L)
3 h:	≥140 mg/dL (≥7.8 mmol/L)	>145 mg/dl (8.0 mmol/L)

If one value is abnormal, repeat test in 4 wk

One step protocol proposed by International Association of the Diabetes and Pregnancy Study Groups:

Fasting glucose at first prenatal visit. Value >92 mg/dL (5.1 mmol/L) is diagnostic of GDM and no additional testing is indicated. If negative then perform diagnostic 75 g OGTT testing at 24 to 28 wk after overnight fast of 8 h. The test is positive if one of the following glucose levels is found.

Fasting:	>92 mg/dL (5.1 mmol/L)
1 h:	>180 mg/dL (10.0 mmol/L)
2 h:	>153 mg/dL (8.5 mmol/L)

WHO Diagnostic criteria and classification of hyperglycemia first detected in pregnancy 2013.

IADPSG Consensus Panel Diabetes Care. 2010;33:676.

National Diabetes Data Group Diabetes. 1979;28:1039.

Carpenter MW, Coustan DR. Am J obstet Gynecol. 1982;144:768.

plasma glucose value of 140 mg/dL (7.8 mmol/L) or greater identifies approximately 80% of women with GDM; decreasing the cutoff value to 130 mg/dL (7.2 mmol/L) increases the sensitivity to approximately 90%. Women with the positive glucose challenge test should undergo a 3-hour 100 g oral glucose tolerance test within the week.

Although there is continuing debate regarding universal screening, recent studies are supportive of this position. An observational study of Caucasian women, where all were tested with the oral glucose tolerance test, found that the prevalence of GDM among women with no risk factors was 2.8%. Excluding this low-risk group would still require 80% of women to be tested and would miss 10% of all cases of GDM. The Hyperglycemia and Pregnancy Outcome (HAPO) study reported a continuous relationship between glucose concentration with the 28-week oral glucose tolerance test and neonatal adiposity, cesarean sections, neonatal hypoglycemia, premature delivery, shoulder dystocia, need for intensive neonatal care, hyperbilirubinemia, and preeclampsia. A study by Landon and colleagues of women with mild GDM on a 100 g 3-hour glucose tolerance test found that treatment reduced the risk of fetal overgrowth, shoulder dystocia, cesarean delivery, and hypertensive disorders. The United States Preventive Services Task Force on Preventive Health Care has concluded that there is good evidence to support universal screening for GDM after 24 weeks but not for universal screening earlier in pregnancy. Women who have a history of GDM with previous pregnancy should undergo preconception counseling similar to those patients with pregestational diabetes. They should begin to monitor their glucose levels when they begin to try to conceive or as soon as their new pregnancy is confirmed. These women do not need retesting for GDM.

HbA$_{1c}$ values are helpful for assessment of pregestational diabetes but are not helpful for management of GDM where they are normal.

The goal of therapy is prevention of fasting and postprandial hyperglycemia. The Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) trial found that intervention to lower glucose levels to less than 100 mg/dL before meals and less than 126 mg/dL 2 hours after meals reduced the rate of serious perinatal outcomes (defined as death, shoulder dystocia, bone fracture, and nerve palsy) from 4% to 1%. After diagnosis, the patient should be placed on a diabetic meal plan modified for pregnancy. The caloric intake is based on ideal body weight—25 to 35 kcal/kg ideal weight, 40% to 55% carbohydrate, 20% protein, and 25% to 40% fat. Calories are distributed over three meals and three snacks (see Table 17–24). Most patients can be taught to count their carbohydrates and to read food labels. This caloric distribution will help 75% to 80% of patients to become normoglycemic. Patients should also be encouraged to participate in moderate aerobic exercise such as walking or antenatal exercise classes, of at least 15- to 30-minutes duration, three or more times a week. In normal pregnancy, expected weight gain varies according to the prepregnancy weight. The Fifth International Workshop-Conference on GDM recommends a relatively small gain during pregnancy of 7 kg (15 lb) for obese women (BMI ≥30 kg/m^2) and a proportionally greater weight gain (up to 18 kg or 40 lb) for underweight women (BMI <18.5 kg/m^2) at the onset of pregnancy. However, there are no data on optimal weight gain for women with GDM.

If fasting capillary blood glucose levels exceed 90 to 100 mg/dL (5-5.6 mmol/L) or if 1-hour or 2-hour postprandial glucose values are consistently greater, respectively, than 130 or 105 mg/dL (7.2 or 5.8 mmol/L) therapy is begun with insulin. The total insulin dose varies from 0.7 to 2 U/kg. NPH or insulin detemir are the preferred long-acting insulins. Insulin glargine has not been well studied. Insulin aspart or lispro and regular insulin can be used. Hypoglycemia is a risk factor with aggressive management of glucose levels with insulin. Patients and family members should be instructed on monitoring for and treating hypoglycemia.

Metformin and glyburide have been considered as alternative options to insulin therapy. Metformin does cross the placenta but has not been associated with teratogenesis. The advantage of this drug is that it does not result in hypoglycemia. An open label study (the MiG trial) of 751 women with GDM at 20 to 33 weeks of gestation randomized to either metformin or insulin did not show a difference in primary composite outcome of neonatal complications. Forty-six percent of the subjects on metformin did, however, require supplemental insulin. Langer and colleagues randomized 404 women with GDM at 11 to 33 weeks of gestation to glyburide therapy or insulin treatment. There was no significant difference in fetal outcomes in the two groups—large for gestational infants, macrosomia, neonatal ICU admission, or fetal abnormalities. Target glycemic control was achieved in 88% of patients on insulin and 82% on glyburide. There was significant reduction in maternal hypoglycemic episodes in the glyburide group compared to the insulin group (2% vs 20%). A large retrospective observational study, however, has raised concerns about glyburide's safety in gestational diabetes. The study reviewed the fetal outcomes of approximately 9000 women with gestational diabetes treated with either glyburide or insulin. They found that infants of women treated with glyburide had higher risk of respiratory distress, neonatal hypoglycemia, birth injury, and large for gestational age.

Patients are usually scheduled for follow-up visits every 1 to 2 weeks. A 24-hour urine collection may be performed to establish baseline level of proteinuria and creatinine clearance due to the higher likelihood of preeclampsia. It is not necessary to routinely perform ophthalmic examinations in these patients unless there is a strong suspicion of preexisting type 2 diabetes mellitus. Fetal monitoring is warranted in those patients who are not well controlled, requiring insulin therapy, or who have other complications of pregnancy. The most commonly used test is a twice-weekly nonstress test.

If blood glucose levels are close to normal and there are no other complications, the delivery can go to term. It is generally recommended that the pregnancy does not go beyond term. Most patients do not require insulin during delivery. Blood glucose levels should be monitored the following day to ensure that the patient has reverted to normoglycemia. About 95% of patients return to normal glucose status.

Progression to type 2 diabetes later in life occurs in 5% to 50% of women with gestational diabetes. The wide range in incidence is influenced by body weight, family history, glucose levels, and

the need for insulin treatment during pregnancy—and the choice of contraception and lifestyle after pregnancy. All patients with gestational diabetes should undergo a 75 g 2-hour glucose tolerance test at 6 to 10 weeks after delivery to guide future medical management. Follow-up protocols after pregnancy and criteria for the diagnosis of diabetes mellitus in the nonpregnant state are presented in Table 17–11.

REFERENCES

The Endocrine Pancreas

Andralojc KM, Mercalli A, Nowak KW, et al. Ghrelin-producing epsilon cells in the developing and adult human pancreas. *Diabetologia*. 2009;52:486. [PMID: 19096824]

Baggio LL, Drucker DJ. Biology of incretins: GLP-1 and GIP. *Gastroenterology*. 2007;132:2131. [PMID: 17498508]

Cabrera O, Berman DM, Kenyon NS, et al. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci USA*. 2006;103:2334. [PMID: 16461897]

Gross DN, van den Heuvel AP, Birnbaum MJ. The role of FOXO in the regulation of metabolism. *Oncogene*. 2008;27:2320. [PMID: 18391974]

Hiriart M, Aguilar-Bryan L. Channel regulation of glucose sensing in the pancreatic beta-cell. *Am J Physiol Endocrinol Metab*. 2008;295:E1298. [PMID: 18940941]

Kieffer TJ, McIntosh CH, Pederson RA. Degradation of glucose-dependent insulinotropic polypeptide and truncated glucagon-like peptide 1 *in vitro* and *in vivo* by dipeptidyl peptidase IV. *Endocrinology*. 1995;136:3585. [PMID: 7628397]

Meier JJ, Butler AE, Saisho Y, et al. Beta-cell replication is the primary mechanism subserving the postnatal expansion of beta-cell mass in humans. *Diabetes*. 2008;57:1584. [PMID: 18334605]

Ohneda K, Ee H, German M. Regulation of insulin gene transcription. *Semin Cell Dev Biol*. 2000;11:227. [PMID: 10966856]

Oliver-Krasinski JM, Stoffers DA. On the origin of the beta cell. *Genes Dev*. 2008;22:1998. [PMID: 18676806]

Rorsman P, Renstrom E. Insulin granule dynamics in pancreatic beta cells. *Diabetologia*. 2003;46:1029. [PMID: 12879249]

Scheepers A, Joost HG, Schürmann A. The glucose transporter families SGLT and GLUT: molecular basis of normal and aberrant function. *JPEN J Parenter Enteral Nutr*. 2004;28:364. [PMID: 15449578]

Sinclair EM, Drucker DJ. Proglucagon-derived peptides: mechanisms of action and therapeutic potential. *Physiology (Bethesda)*. 2005;20:357. [PMID: 16174875]

Steinberg GR, Kemp BE. AMPK in health and disease. *Physiol Rev*. 2009;89:1025. [PMID: 19584320]

Tontonoz P, Spiegelman BM. Fat and beyond: the diverse biology of PPAR gamma. *Annu Rev Biochem*. 2008;77:289. [PMID: 18518822]

Vaxillaire M, Froguel P. Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. *Endocr Rev*. 2008;29:254. [PMID: 18436708]

Wilson ME, Scheel D, German MS. Gene expression cascades in pancreatic development. *Mech Dev*. 2003;120:65. [PMID: 12490297]

Diagnosis, Classification, and Pathophysiology of Diabetes Mellitus

Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358:2545. [PMID: 18539917]

ADVANCE Collaborative Group. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2008;358:2560. [PMID: 18539916]

American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2014;37(Suppl 1):S81. [PMID: 24357215]

Ashcroft FM, Rorsman P. Diabetes mellitus and the β cell: the last ten years. *Cell*. 2012;148(6):1160. [PMID: 2242422]

Biddinger SB, Kahn CR. From mice to men: insights into the insulin resistance syndromes. *Annu Rev Physiol*. 2006;68:123. [PMID: 16460269]

Bluestone JA, Herold K, Eisenbarth G. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature*. 2010;464:1293. [PMID: 20432533]

Brown RJ, Araujo-Vilar D, Cheung PT, et al. The Diagnosis and Management of Lipodystrophy Syndromes: A Multi-Society Practice Guideline. *J Clin Endocrinol Metab*. 2016;101:4500. [PMID: 27710244]

Butler AE, Janson J, Bonner-Weir S, et al. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes*. 2003;52:102. [PMID: 12502499]

Cohen RM, Holmes YR, Chenier TC, Joiner CH. Discordance between HbA1c and fructosamine: evidence for a glycosylation gap and its relation to diabetic nephropathy. *Diab Care*. 2003;26:16. [PMID: 12502674]

Diabetes Prevention Trial, Type 1 Diabetes Study Group. Effects of insulin in relatives of patients with type 1 diabetes mellitus. *N Engl J Med*. 2002;346:1685. [PMID: 12037147]

Duckworth W, Abraira C, Moritz T, et al. VADT Investigators. Glucose control and vascular complications in veterans with type 2 diabetes. *N Engl J Med*. 2009;360:129. [PMID: 19092145]

Flanagan SE, De Franco E, Lango Allen H, et al. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. *Cell Metab*. 2014;19:146. [PMID: 24411943]

Fonseca SG, Fukuma M, Lipson KL, et al. WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic beta-cells. *J Biol Chem*. 2005;280:39609. [PMID: 16195229]

Gaede P, Lund-Andersen H, Parving HH, Pedersen O. Effect of a multifactorial intervention on mortality in type 2 diabetes. *N Engl J Med*. 2008;358:580. [PMID: 18256393]

Gaede P, Vedel P, Larsen N, Jensen GV, Parving HH, Pedersen O. Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes. *N Engl J Med*. 2003;348:383. [PMID: 12556541]

Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359:1577. [PMID: 18784090]

Holman RR, Paul SK, Bethel MA, Neil HA, Matthews DR. Long-term follow-up after tight control of blood pressure in type 2 diabetes. *N Engl J Med*. 2008;359:1565. [PMID: 18784091]

International Expert Committee. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diab Care*. 2009;32:1327 [PMID: 19502545]

Johnson MB, Hattersley AT, Flanagan SE. Monogenic autoimmune diseases of the endocrine system. *Lancet Diabetes Endocrinol*. 2016;4:862. [PMID: 27474216]

Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002;346:393. [PMID: 11832527]

Kahn R, Buse J, Ferrannini E, et al. The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diab Care*. 2005;28:2289. [PMID: 16123508]

Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ. A1c-Derived Average Glucose Study Group. Translating the A1C assay into estimated average glucose values. *Diab Care*. 2008;31(8):1473. [PMID: 18540046]

Odegaard JI, Chawla A. Pleiotropic actions of insulin resistance and inflammation in metabolic homeostasis. *Science*. 2013;339(6116):172. [PMID: 23307735]

Ohkubo Y, Kishikawa H, Araki E, et al. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. *Diab Res Clin Pract*. 1995;28:103. [PMID: 7587918]

Okamoto M, Okamoto M, Gotoh K, et al. Fulminant type 1 diabetes mellitus with anti-programmed cell death-1 therapy. *J Diabetes Investig*. 2016;7:915. [PMID: 27181090]

Schwitzgebel VM. Many faces of monogenic diabetes. *J Diabetes Investig*. 2014;5:121. [PMID: 24843749]

Selph S, Dana T, Bougatsos C, et al. Screening for Abnormal Glucose and Type 2 Diabetes Mellitus: A Systematic Review to Update the 2008 U.S. Preventive Services Task Force. *Ann Intern Med*. 2015;162:765. [PMID: 25867111]

Shichiri M, Kishikawa H, Ohkubo Y, Wake N. Long term results of the Kumamoto Study on optimal diabetes control in type 2 diabetic patients. *Diab Care*. 2000;23(Suppl 2):B21. [PMID: 10860187]

Stekelenburg CM, Schwitzgebel VM. Genetic defects of the beta-cell that cause diabetes. *Endocr Dev*. 2016;31:179. [PMID: 26824366]

Stumvoll M, Gerich J. Clinical features of insulin resistance and beta cell dysfunction and the relationship to type 2 diabetes. *Clin Lab Med*. 2001;21:31. [PMID: 11321936]

The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993;329:977. [PMID: 8366922]

The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. *N Engl J Med*. 2000;342:381. [PMID: 10666428]

The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*. 2005;353:2643. [PMID: 16371630]

The Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N Engl J Med*. 2008;359:1464. [PMID: 18779236]

Thomas CC, Philipson LH. Update on diabetes classification. *Med Clin North Am*. 2015;99:1. [PMID: 25456640]

Turnbull FM, Abraira C, Anderson RJ, et al. Intensive glucose control and macrovascular outcomes in type 2 diabetes. *Diabetologia*. 2009;52:2288. [PMID: 19655124]

Turner RC, Cull CA, Frighi V, Holman RR. Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). *JAMA*. 1999;281:2005. [PMID: 10359389]

UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet*. 1998;352:854.

UK Prospective Diabetes Study Group. Efficacy of atenolol and captopril in reducing risk of acrovascular and microvascular complications in type 2 diabetes: UKPDS 39. *BMJ*. 1998;317:713. [PMID: 9732338]

UK Prospective Diabetes Study Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet*. 1998;352:837. [PMID: 9742976]

UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes (UKPDS38). *BMJ*. 1998;317:703. [PMID: 732337]

Vaxillaire M, Froguel P. Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. *Endocr Rev*. 2008;29:254. [PMID: 18436708]

Wajchenberg BL. Subcutaneous and visceral adipose tissue: their relation to the metabolic syndrome. *Endocr Rev*. 2000;21:697. [PMID: 11133069]

Wenzlau JM, Frisch LM, Gardner TJ, et al. Novel antigens in type 1 diabetes: the importance of ZNT8. *Curr Diab Rep*. 2009;9:105. [PMID: 19323954]

Treatment of Diabetes Mellitus

American Diabetes Association. Standards of medical care in diabetes—2017. *Diab Care*. 2016;40:S1.

Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014;383:69. [PMID: 23890997]

Bays HE, Jones PH, Orringer CE, et al. National Lipid Association Annual Summary of Clinical Lipidology 2016. *J Clin Lipidol*. 2016;10(Suppl 1):S1. [PMID: 26891998]

Brunkhorst FM, Engel C, Bloos F, et al. German Competence Network Sepsis (SepNet). Intensive insulin therapy and pentastarch resuscitation in severe sepsis. *N Engl J Med*. 2008;358:125. [PMID: 18184958]

Brunström M, Carlberg B. Effect of antihypertensive treatment at different blood pressure levels in patients with diabetes mellitus: systematic review and meta-analyses. *BMJ*. 2016;352:i717. [PMID: 26920333]

Coursin DB, Connery LE, Ketzler JT. Perioperative diabetic and hyperglycemic management issues. *Crit Care Med*. 2004;32(4 Suppl):S116. [PMID: 15064670]

Cryer PE. The barrier of hypoglycemia in diabetes. *Diabetes*. 2008;57:3169. [PMID: 19033403]

Daneman D. Type 1 diabetes. *Lancet*. 2006;367(9513):847. [PMID: 16530579]

Del Pilar Solano M, Goldberg RB. Management of diabetic dyslipidemia. *Endocrinol Metab Clin North Am*. 2005;34:1. [PMID: 15752919]

DeWitt DE, Hirsch IB. Outpatient insulin therapy in type 1 and type 2 diabetes mellitus: scientific review. *JAMA*. 2003;289:2254. [PMID: 12734137]

Emdin CA, Rahimi K, Neal B, et al. Blood pressure lowering in type 2 diabetes: a systematic review and meta-analysis. *JAMA*. 2015;313:603. [PMID: 25668264]

Evert AB, Boucher JL, Cypress M, et al. Nutrition therapy recommendations for the management of adults with diabetes. *American Diabetes Association*. *Diabetes Care*. 2013;36:3821. [PMID: 24107659]

Gururaj Setty S, Crasto W, Jarvis J, et al. New insulins and newer insulin regimens: a review of their role in improving glycaemic control in patients with diabetes. *Postgrad Med J*. 2016;92:152. [PMID: 26842973]

Harlan DM. Islet transplantation for hypoglycemia unawareness/severe hypoglycemia: caveat emptor. *Diabetes Care*. 2016;39:1072. [PMID: 27330121]

Inzucchi SE, Bergenstal RM, Buse JB, et al. Management of hyperglycaemia in type 2 diabetes, 2015: a patient-centered approach. Update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetologia*. 2015;58:429. [PMID: 25583541]

Kakkar AK, Dahiya N. Drug treatment of obesity: current status and future prospects. *Eur J Intern Med*. 2015;26:89. [PMID: 25634851]

Khan MA, St Peter JV, Xue JL. The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diab Care*. 2005;28:2289. [PMID: 16123508]

Khera R, Murad MH, Chandar AK, et al. Association of pharmacological treatments for obesity with weight loss and adverse events: a systematic review and meta-analysis. *JAMA*. 2016;315:2424. [PMID: 2729961]

Kropff J, DeVries JH. Continuous Glucose Monitoring, Future Products, and Update on Worldwide Artificial Pancreas Projects. *Diabetes Technol Ther*. 2016;18(Suppl 2):S253. [PMID: 2678413]

Lee C-H, Olson P, Evans RM. Lipid metabolism, metabolic diseases, and peroxisome proliferator-activated receptors. *Endocrinology*. 2003;144:2201. [PMID: 12746275]

Massicotte A. Contrast medium-induced nephropathy: strategies for prevention. *Pharmacother*. 2008;28:1140. [PMID: 18752385]

Moghissi ES, Korytkowski MT, DiNardo M, et al. American Association of Clinical Endocrinologists and American Diabetes Association consensus statement on inpatient glycemic control. *Diabetes Care*. 2009;32(6):1119. [PMID: 19429873]

Nauck M. Incretin therapies: highlighting common features and differences in the modes of action of glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors. *Diabetes Obes Metab*. 2016;18:203. [PMID: 26489970]

NICE-SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*. 2009;360:1283. [PMID: 19318384]

Niclauss N, Meier R, Bédat B, et al. Beta-cell replacement: pancreas and islet cell transplantation. *Endocr Dev*. 2016;31:146. [PMID: 26824893]

Pickup JC. Management of diabetes mellitus: is the pump mightier than the pen? *Nat Rev Endocrinol*. 2012;8:425. [PMID: 22371161]

Taskinen MR, Borén J. New insights into the pathophysiology of dyslipidemia in type 2 diabetes. *Atherosclerosis*. 2015;239:483. [PMID: 25706066]

Thabit H, Hovorka R. Coming of age: the artificial pancreas for type 1 diabetes. *Diabetologia*. 2016;59:1795. [PMID: 27364997]

Wright JT Jr, Williamson JD, Whelton PK, et al. SPRINT Research Group, A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med*. 2015;373:2103 [PMID: 26551272]

Zinman B, Ruderman N, Campaigne BN, Devlin JT, Schneider SH. American Diabetes Association. Physical activity/exercise and diabetes. *Diabetes Care*. 2004;27(Suppl 1):S58. [PMID: 14693927]

Zoungas S, Ninomiya T, Huxley R, et al. Systematic review: sodium bicarbonate treatment regimens for the prevention of contrast-induced nephropathy. *Ann Intern Med*. 2009;151:631. [PMID: 19884624]

Acute Complications of Diabetes Mellitus

Henderson G. The psychosocial treatment of recurrent diabetic ketoacidosis: an interdisciplinary team approach. *Diabetes Educator*. 1991;17:119. [PMID: 1899827]

Kitabchi AE, Nyenwe EA. Hyperglycemic crises in diabetes mellitus: diabetic ketoacidosis and hyperglycemic hyperosmolar state. *Endocrinol Metab Clin North Am*. 2006;35:725. [PMID: 17127143]

Luft FC. Lactic acidosis update for critical care clinicians. *J Am Soc Nephrol*. 2001;12(Suppl 17):S15. [PMID: 11251027]

Nugent BW. Hyperosmolar hyperglycemic state. *Emerg Med Clin North Am*. 2005;23:629. [PMID: 15982538]

Okuda Y, Adrogue HJ, Field JB, Nohara H, Yamashita K. Counterproductive effects of sodium bicarbonate in diabetic ketoacidosis. *J Clin Endocrinol Metab*. 1996;81:314. [PMID: 8550770]

Reno CM, Litvin M, Clark AL, Fisher SJ. Defective counterregulation and hypoglycemia unawareness in diabetes: mechanisms and emerging treatments. *Endocrinol Metab Clin North Am*. 2013;42:15. [PMID: 23391237]

Salpeter SR, Greyber E, Pasternak GA, Salpeter EE. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. *Cochrane Database Syst Rev*. 2010;4:CD002967. [PMID: 20091535]

Seaquist ER, Anderson J, Childs B, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *J Clin Endocrinol Metab*. 2013;98(5):1845. [PMID: 23589524]

Wrenn KD, Slovis CM, Minion GE, Rutkowski R. The syndrome of alcoholic ketoacidosis. *Am J Med*. 1991;91:119. [PMID: 1867237]

Chronic Complications of Diabetes Mellitus

American Diabetes Association Position Statement. Preventive foot care in people with diabetes. *Diab Care*. 2004;27(Suppl 1):S63.

Bowling FL, Rashid ST, Boulton AJ. Preventing and treating foot complications associated with diabetes mellitus. *Nat Rev Endocrinol*. 2015;11:606. [PMID: 26284447]

Boulton AJ. The diabetic foot: from art to science. The 18th Camillo Golgi lecture. *Diabetologia*. 2004;47:1343. [PMID: 15309286]

Boulton AJ, Vinik AI, Arezzo JC, et al. Diabetic neuropathies: a statement by the American Diabetes Association. *Diab Care*. 2005;28:956. [PMID: 15793206]

Brownlee M. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes*. 2005;54:1615. [PMID: 15919781]

Camilleri M, Parkman HP, Shafi MA, et al. Clinical guideline: management of gastroparesis. American College of Gastroenterology. *Am J Gastroenterol*. 2013;108(1):18. [PMID: 23147521]

Deckert T, Kofoed-Enevoldsen A, Norgaard K, et al. Microalbuminuria: implications for micro- and macrovascular disease. *Diab Care*. 1992;15:1181. [PMID: 1396015]

Fineberg D, Jandeleit-Dahm KA, Cooper ME. Diabetic nephropathy: diagnosis and treatment. *Nat Rev Endocrinol*. 2013;9:713. [PMID: 2410026]

Mascarenhas JV, Albayati MA, Shearman CP, et al. Peripheral arterial disease. *Endocrinol Metab Clin North Am*. 2014;43:149. [PMID: 24582096]

Mauer SM, Goetz FC, McHugh LE, et al. Long-term study of normal kidneys transplanted into patients with type I diabetes. *Diabetes*. 1989;38:516. [PMID: 2647558]

Stitt AW, Curtis TM, Chen M, et al. The progress in understanding and treatment of diabetic retinopathy. *Prog Retin Eye Res*. 2016;51:156. [PMID: 26297071]

Tesfaye S, Vileikyte L, Rayman G, et al. Painful diabetic peripheral neuropathy: consensus recommendations on diagnosis, assessment and management. Toronto Expert Panel on Diabetic Neuropathy. *Diabetes Metab Res Rev*. 2011;27:629. [PMID: 21695762]

Thazhath SS, Jones KL, Horowitz M, et al. Diabetic gastroparesis: recent insights into pathophysiology and implications for management. *Expert Rev Gastroenterol Hepatol*. 2013;7:127. [PMID: 23363262]

Vardi M, Nini A. Phosphodiesterase inhibitors for erectile dysfunction in patients with diabetes mellitus. *Cochrane Database Syst Rev*. 2007;1:CD002187. [PMID: 17253475]

Diabetes Mellitus and Pregnancy

Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med*. 2005;352:2477. [PMID: 15951574]

Boinpally T, Jovanovic L. Management of type 2 diabetes and gestational diabetes in pregnancy. *Mt Sinai J Med*. 2009;76:269. [PMID: 19421970]

HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*. 2008;358:1991. [PMID: 18463375]

Kim H, Toyofuku Y, Lynn FC, et al. Serotonin regulates pancreatic beta cell mass during pregnancy. *Nat Med*. 2010;16:804. [PMID: 20581837]

Kitzmiller JL, Block JM, Brown FM, et al. Managing preexisting diabetes for pregnancy: summary of evidence and consensus recommendations for care. *Diab Care*. 2008;31:1060. [PMID: 18445730]

Kitzmiller JL, Gavin LA, Gin GD, et al. Preconception management of diabetes continued through early pregnancy prevents the excess frequency of major congenital anomalies in infants of diabetic mothers. *JAMA*. 1991;265:731. [PMID: 1990188]

Landon MB, Spong CY, Thom E, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med*. 2009;361:1339. [PMID: 19797280]

Langer O, Conway DL, Berkus MD, Xenakis EM, Gonzales O. A comparison of glyburide and insulin in women with gestational diabetes. *N Engl J Med*. 2000;343:1134.

MiG Trial Investigators. Metformin versus insulin for the treatment of gestational diabetes. *N Engl J Med*. 2008;358:2003. [PMID: 18463376]

Moyer VA; U.S. Preventive Services Task Force. Screening for gestational diabetes mellitus: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160:414. [PMID: 24424622]

Murphy HR, Rayman G, Lewis K, et al. Effectiveness of continuous glucose monitoring in pregnant women with diabetes: randomised clinical trial. *BMJ*. 2008;337:1680. [PMID: 18818254]

Walkinshaw SA. Pregnancy in women with pre-existing diabetes: management issues. *Semin Fetal Neonatal Med*. 2005;10:307. [PMID: 15927547]

World Health Organization; 2013. Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy. [PMID: 24199271]

Hypoglycemic Disorders

Umesh Masharani, MB, BS, MRCP (UK),
Stephen E. Gitelman, MD, and Roger K. Long, MD

ACE	Angiotensin-converting enzyme		IGFBP3	IGF-binding protein-3
ACTH	Adrenocorticotrophic hormone		K_{ATP}	ATP-sensitive potassium channel
ADP	Adenosine diphosphate		Kir	Inward rectifier potassium channel
ALS	Acid-labile subunit		NADH	Nicotinamide adenine dinucleotide (reduced form)
ATP	Adenosine triphosphate			
GDH	Glutamate dehydrogenase		NICTH	Nonislet cell tumor hypoglycemia
GIP	Glucose-dependent insulinotropic polypeptide		NIPHS	Noninsulinoma pancreatogenous hypoglycemic syndrome
GLP-1	Glucagon-like peptide 1			
G-6-Pase	Glucose-6-phosphatase		PEPCK	Phosphoenolpyruvate carboxykinase
HbA_{1c}	Hemoglobin A_{1c}		RIA	Radioimmunoassay
ICMA	Immunochemiluminometric		SSTR	Somatostatin receptor
IGF-II	Insulin-like growth factor II		SUR	Sulfonylurea receptor

Circulating plasma glucose concentrations are kept within a relatively narrow range by a complex system of interrelated neural, humoral, and cellular controls. Under the usual metabolic conditions, the central nervous system is wholly dependent on plasma glucose and counteracts declining blood glucose concentrations with a carefully programmed response. This is often associated with a sensation of hunger and, as the brain receives insufficient glucose to meet its metabolic needs (neuroglycopenia), an autonomic response is triggered to mobilize storage depots of glycogen and fat. In the postabsorptive state, hepatic glycogen reserves and gluconeogenesis from the liver and kidney directly supply the central nervous system with glucose, which is carried across the blood–brain barrier by a specific glucose transport system, while the mobilization of fatty acids from triglyceride depots provides energy for the large mass of skeletal and cardiac muscle, renal cortex, liver, and other tissues that utilize fatty acids as their basic fuel, thus sparing glucose for use by the tissues of the central nervous system.

The normal lower limit of fasting plasma glucose is typically 70 mg/dL (3.9 mmol/L). Lower values may occur during prolonged fasting, strenuous exercise, or pregnancy or may occur as a laboratory artifact. In normal men, plasma glucose does not fall below 55 mg/dL (3 mmol/L) during a 72-hour fast. However, for reasons that are not clear, normal women may experience a fall to levels as low as 30 mg/dL (1.7 mmol/L) despite a marked suppression of circulating insulin to less than 5 μU/mL. They remain asymptomatic in spite of this degree of hypoglycemia, presumably because ketogenesis is able to satisfy the energy needs of the central nervous system. Basal plasma glucose declines progressively during normal pregnancy, and hypoglycemic levels may be reached during prolonged fasting. This may be a consequence of a continuous fetal consumption of glucose and diminished availability of the gluconeogenic substrate alanine. The cause of these diminished alanine levels in pregnancy is unclear. The greatly increased glucose consumption by skeletal muscle that occurs during prolonged strenuous exercise may lead to hypoglycemia (blood glucose <45 mg/dL) despite increases in hepatic glucose production. Whether the hypoglycemia in this circumstance contributes to fatigue or other symptoms in distance runners is unknown.

In vitro consumption of glucose by blood cell elements has been reported to average 10 mg/dL/h and may give rise to laboratory values in the hypoglycemic range. This can be avoided by adding a small amount of the metabolic inhibitor sodium fluoride to collection tubes used for specimens containing increased numbers of blood cells (as in leukemia, leukemoid reactions, or polycythemia) and keeping samples on ice until separated.

PATHOPHYSIOLOGY OF THE COUNTERREGULATORY RESPONSE TO NEUROGLYCOPENIA

The plasma concentration of glucose that signals the need by the central nervous system to mobilize energy reserves depends on a number of factors, such as the status of blood flow to the brain, the integrity of cerebral tissue, the prevailing arterial level of plasma glucose, the rapidity with which plasma glucose concentration falls, and the availability of alternative metabolic fuels.

A hierarchy of responses has been shown to occur as plasma glucose falls in healthy young volunteers, with hormonal counterregulatory responses being triggered at glucose levels slightly higher (~67 mg/dL [3.7 mmol/L]) than those that induce symptoms of hypoglycemia (Figure 18–1). The first symptoms to appear in healthy people are mediated by autonomic neurotransmitters and occur at plasma glucose levels below 60 mg/dL (3.3 mmol/L). The symptoms consist of tremor, anxiety, palpitations, and sweating, which result from sympathetic discharge; and hunger, which is a consequence of parasympathetic vagal response. Ganglionic blockade and cervical cord section or sympathectomy—but not adrenalectomy—ameliorates these symptoms, indicating that they are due to the release of autonomic neurotransmitters and not dependent on adrenal hormones. As plasma glucose falls below 50 mg/dL (2.8 mmol/L), cerebral neuroglycopenia ensues, consisting of impaired cognition, along with weakness, lethargy, confusion, incoordination, and blurred vision. If counterregulatory responses are inadequate to reverse this degree of profound hypoglycemia, convulsions or coma may occur. This can result in brain damage or death, particularly in those who have not adapted to repeated episodes of hypoglycemia.
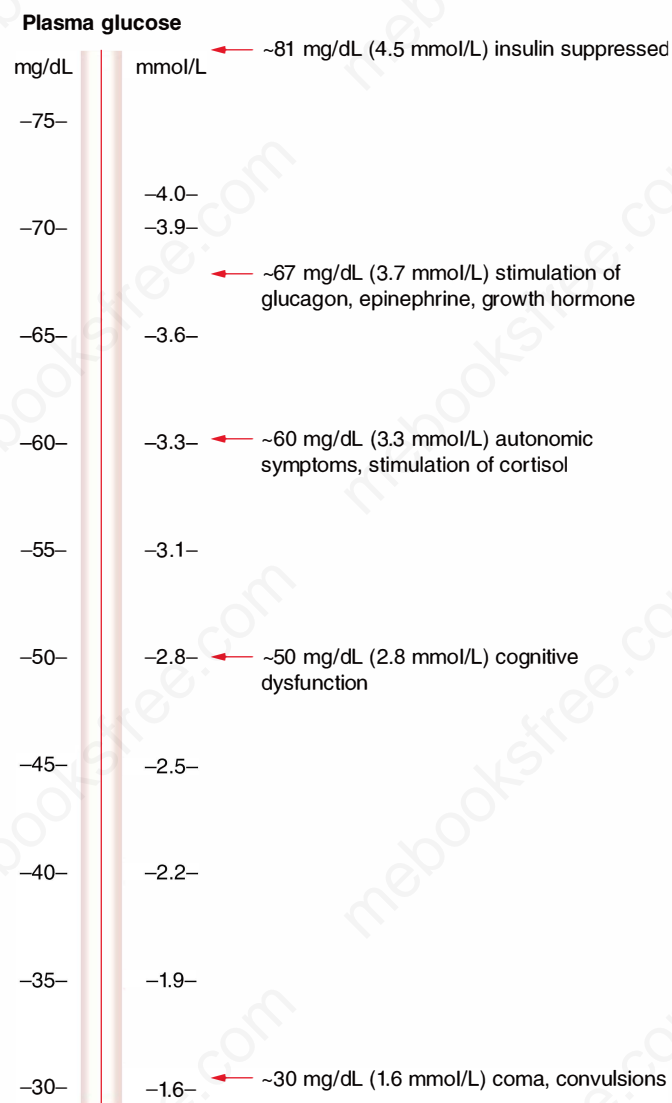
In elderly people, however, with compromised cerebral blood supply, neuroglycopenic manifestations may be provoked at slightly higher plasma glucose levels. Patients with chronic hyperglycemia (eg, those with poorly controlled, insulin-treated diabetes mellitus) may experience symptoms of neuroglycopenia at considerably higher plasma glucose concentrations than persons without diabetes. This has been attributed to a down-regulated glucose transport system across the blood–brain barrier. Conversely, in patients exposed to chronic hypoglycemia (eg, those with an insulin-secreting tumor or those with diabetes who are receiving excessively tight glycemic control with insulin), adaptation to recurrent hypoglycemia occurs by upregulation of the glucose transporters, which results in hypoglycemic unawareness, whereby they show greater tolerance to hypoglycemia without manifesting symptoms. Restoring and maintaining an adequate supply of glucose for cerebral function proceeds by a series of neurogenic events that act directly to raise the plasma glucose concentration and to stimulate hormonal responses that augment the adrenergic mobilization of energy stores (Table 18–1).

FIGURE 18–1 Hierarchy of hormonal and neurological responses to progressive stepwise reduction in plasma glucose concentration in healthy volunteers.

TABLE 18–1 Autonomic nervous system response to hypoglycemia.

Alpha-Adrenergic Effects
Inhibition of endogenous insulin release
Increase in cerebral blood flow (peripheral vasoconstriction)

Beta-Adrenergic Effects
Hepatic and muscle glycogenolysis
Stimulation of plasma glucagon release
Lipolysis to raise plasma-free fatty acids
Impairment of glucose uptake by muscle tissue
Increase in cerebral blood flow (increase in cardiac output)

Adrenomedullary Discharge of Catecholamines
Augmentation of all of the above alpha- and beta-adrenergic effects

Cholinergic Effects
Raises level of pancreatic polypeptide
Increases motility of stomach
Produces hunger
Increases sweating

Counterregulatory Response to Hypoglycemia (Figures 18–2 and 18–3)

Insulin   Endogenous insulin secretion is lowered both by reduced glucose stimulation to the pancreatic β cell and by sympathetic nervous system inhibition from a combination of alpha-adrenergic neural effects and increased circulating catecholamine levels. This reactive insulinopenia appears to be essential for glucose recovery, because it facilitates the mobilization of energy from existing energy stores (glycogenolysis and lipolysis); increases hepatic enzymes involved in gluconeogenesis and ketogenesis; increases enzymes of the renal cortex, promoting gluconeogenesis; and at the same time prevents muscle tissue from consuming the blood glucose being released from the liver (see Chapter 17).

Catecholamines   Circulating catecholamines—and norepinephrine produced at sympathetic nerve endings—provide muscle tissue with alternative sources of fuel by activating beta-adrenergic receptors, resulting in mobilization of muscle glycogen, and by providing increased plasma-free fatty acids from lipolysis of adipocyte triglyceride. Metabolism of these free fatty acids provides energy to promote gluconeogenesis in the liver and kidney, thereby adding to plasma glucose levels already raised by the glycogenolytic effect of catecholamines on the liver. Their cardiovascular and other side effects provide a signal that diabetic patients learn to recognize as a warning of their need to ingest rapidly absorbable carbohydrate.

Glucagon   Plasma glucagon is released by the beta-adrenergic effects of both sympathetic innervation and circulating catecholamines on pancreatic α cells as well as by the direct stimulation of α cells by the low plasma glucose concentration. Data are available suggesting that a falling-off of intra-islet insulin concentration in subjects with functioning pancreatic β cells can release pancreatic α cells from insulin inhibition and thus augment glucagon release during hypoglycemia. This glucagon release increases hepatic output of glucose by direct glycogenolysis as well as by facilitating the activity of gluconeogenic enzymes in the liver but not in the kidney. As shown in Figure 18–3, plasma glucagon appears to be the key counterregulatory hormone affecting recovery from *acute* hypoglycemia in nondiabetic humans, with the adrenergic-catecholamine response representing a major backup system. However, in most clinical situations, where hypoglycemia develops *more gradually*, as with inappropriate dosage of insulin or sulfonylureas, or in cases of insulinoma, the role of glucagon may be less influential. When normal volunteers received a prolonged low-dose insulin infusion to produce a gradual decline in plasma glucose levels without waning of insulin levels, the rise of endogenous glucagon contributed much less to counterregulation than after acute hypoglycemia induced by an intravenous insulin bolus, which is followed by rapid waning of insulin levels. This finding suggests that glucagon's role in glucose recovery occurs primarily when the level of insulin wanes.
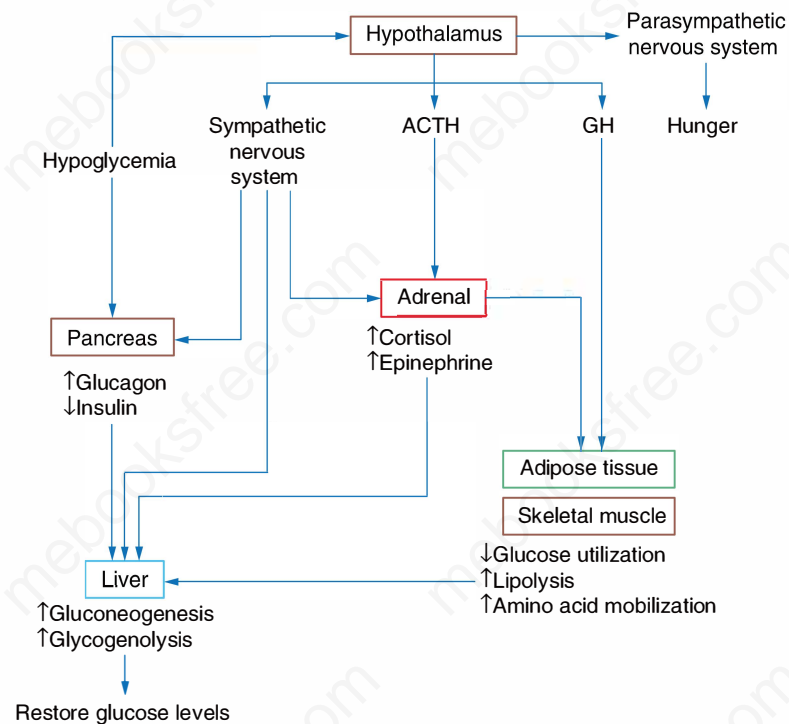
FIGURE 18–2   Regulatory response to hypoglycemia. A fall in glucose level in the brain activates the sympathetic system; induces hunger (parasympathetic system); releases ACTH and GH. The decline in glucose levels and sympathetic activation inhibits insulin secretion and stimulates glucagon from pancreatic islets. Epinephrine and cortisol are released from the adrenals. These hormonal changes and sympathetic nervous system activation act in concert to facilitate glycogenolysis, gluconeogenesis and lipolysis, to limit peripheral glucose utilization and to restore glucose levels to normal.
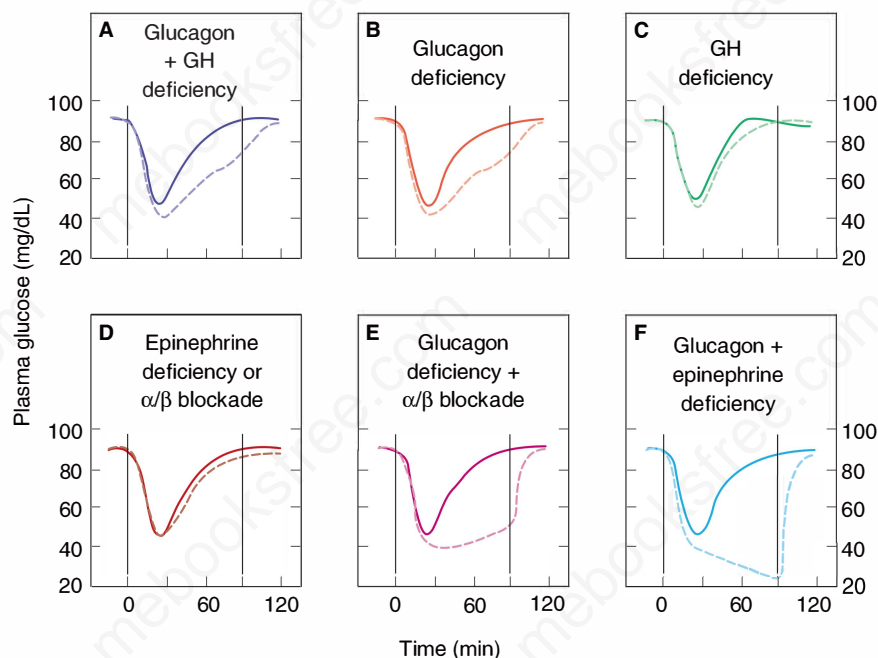
FIGURE 18–3   Solid lines show changes in plasma glucose that occur in normal subjects in response to acute intravenous insulin administration. Note the rapid recovery of glucose levels mediated by intact counterregulatory mechanisms. The dashed lines show the response to insulin-induced hypoglycemia in patients with deficiencies of the counterregulatory mechanisms induced as follows: **A:** Somatostatin infusion (inhibits both glucagon and growth hormone [GH] release). **B:** Somatostatin infusion plus GH infusion (now with functional isolated glucagon deficiency). **C:** Somatostatin infusion plus glucagon infusion (isolated GH deficiency). Note return of glucose response to normal, implying that glucagon is the main counterregulatory hormone. **D:** Bilateral adrenalectomy, leading to epinephrine deficiency, or infusion of phentolamine plus propranolol (alpha and beta blockers, respectively). Note that such deficiencies cause no major abnormality in response to induced hypoglycemia when glucagon is present. **E and F:** Sympathetic modulation (by phentolamine plus propranolol in **E** and by bilateral adrenalectomy in **F**), which seriously impairs the ability to respond to hypoglycemia in the patient made glucagon-deficient by somatostatin infusion. (Data from Clarke WL, Santiago JV, Thomas L, et al. The role of adrenergic mechanisms in recovery from hypoglycemia in man: studies with adrenergic blockade. *Am J Physiol* 236:E147-152, 1979; Gerich J, Davis J, Lorenzi M, et al. Hormonal mechanisms of recovery from insulin-induced hypoglycemia in man. *Am J Physiol* 236:E380-385, 1979; Rizza RA, Cryer PE, Gerich JE. Role of glucagon, epinephrine and growth hormone in human glucose counterregulation: effects of somatostatin and adrenergic blockade on plasma glucose recovery and glucose flux rates following insulin-induced hypoglycemia. *J Clin Invest* 64:62-71, 1979.)

Cortisol   Pituitary corticotrophin (ACTH) is released in association with the sympathetic nervous system stimulation by neuroglycopenia. This results in elevation of plasma cortisol levels, which in turn permissively facilitates lipolysis and actively promotes protein catabolism and conversion of amino acids to glucose by the liver and kidney.

Growth hormone   Pituitary growth hormone is also released in response to falling plasma glucose levels. Its role in counteracting hypoglycemia is less well defined, but it is known to antagonize the action of insulin on glucose utilization in muscle cells and to directly activate lipolysis by adipocytes. This increased lipolysis provides fatty acid substrate to the liver and renal cortex which facilitates gluconeogenesis.

Cholinergic neurotransmitters   Acetylcholine is released at parasympathetic nerve endings, and its vagal effects induce the sensation of hunger that signals the need for food to counteract the hypoglycemia. In addition, postsynaptic fibers of the sympathetic nervous system that innervate the sweat glands to signal hypoglycemia also release acetylcholine—in contrast to all other sympathetic postsynaptic fibers, which without exception release norepinephrine.

Maintenance of Euglycemia in the Postabsorptive State

Glucose absorption from the gastrointestinal tract ceases by 4 to 6 hours after a meal. During the postabsorptive state which immediately follows, glucose must be produced endogenously from previously stored nutrients to meet the requirements of the central nervous system and other glucose-dependent tissues. These include 125 mg/min of glucose required by the brain and spinal cord as well as an additional 25 mg/min by red blood cells and the renal medulla. It was previously thought that the liver is the only organ involved in glucose production during an overnight fast, but recent data indicate that the renal cortex also has the requisite enzymes for production and release of glucose.

Liver   The liver initially provides glucose by the breakdown of stored hepatic glycogen. However, because these reserves are limited to 80 to 100 g, they begin to be depleted several hours into the postabsorptive state. Thereafter, hepatic glucose production is augmented by gluconeogenesis—the formation of glucose from amino acids, lactate, and glycerol. These substrates are delivered to the liver and kidney from peripheral stores. Muscle and other

structural tissues supply amino acids, mainly alanine; blood cell elements supply lactate, the end product of glycolytic metabolism; and adipose tissue supplies glycerol from lipolysis of triglyceride. In addition, oxidation of the free fatty acids released from adipose cells during lipolysis supplies the energy required for gluconeogenesis and provides ketone bodies, acetoacetate, and β-hydroxybutyrate, which can serve as alternative metabolic fuels for the central nervous system during periods of prolonged fasting. Studies have shown that an insulin infusion does not reduce hepatic glucose production if elevated levels of fatty acids are maintained by intravenous administration of a fat emulsion and heparin, suggesting that fatty acids may be the major mediator of gluconeogenesis.

The principal mechanisms regulating hepatic glucose output are the availability of gluconeogenic substrates and the regulation of transcription of key regulatory enzymes such as phosphoenolpyruvate carboxykinase (PEPCK) and glucose-6-phosphatase (G-6-Pase). Glucagon and glucocorticoids through signaling intermediates induce the expression of PEPCK and G-6-Pase. Insulin on the other hand suppresses the expression of these enzymes.

Kidney Although it is generally acknowledged that after fasting for 60 hours the kidney contributes up to 20% to 25% of endogenous glucose production, its role after an overnight fast remains controversial. One group has found that it produces as much as 25% of the postabsorptive glucose requirement, yet a second group using different methodology found a contribution of no more than 5% in the postabsorptive state. Despite these conflicting findings, it is clear that because the renal medulla removes almost as much glucose as the renal cortex produces, the net renal contribution is minimal in the short-term postabsorptive state as compared to the liver.

The kidney does not have glycogen stores and is dependent on gluconeogenesis as its only source of glucose production. Glutamine—rather than alanine—is the predominant amino acid substrate for renal gluconeogenesis. In addition to its contribution to glucose homeostasis after an overnight fast, the kidney also has been shown to be an important contributor to glucose counterregulation in the event of hypoglycemia. Although glucagon does not affect the kidney, the counterregulatory rise in catecholamines has been shown to stimulate gluconeogenesis in the renal cortex. As with the liver, insulin inhibits renal gluconeogenesis and glucose release. Hormonal changes that begin early in the postabsorptive state regulate the enzymatic steps necessary for hepatic glycogenolysis and hepatic and renal gluconeogenesis and ensure the delivery of the necessary substrate (Table 18–2). An appropriate fall in circulating insulin levels with a corresponding rise in glucagon is most important; elevations in the counterregulatory hormones, namely cortisol and growth hormone, contribute to it but are less critical. Thus, numerous endocrine and metabolic events interact to provide a continuous source of fuel for proper functioning of the central nervous system. Malfunction of any of these mechanisms can lead to symptomatic hypoglycemia.

TABLE 18–2 Hormonal changes to maintain euglycemia in the postabsorptive state.

Decreased Insulin Secretion
Increases hepatic glycogenolysis
Increases lipolysis
Increases hepatic gluconeogenesis
Decreases muscle uptake of glucose
Increased Glucagon Secretion
Increases hepatic glycogenolysis
Facilitates hepatic gluconeogenesis
Increased Cortisol Secretion
Facilitates lipolysis
Increases protein catabolism
Augments hepatic gluconeogenesis

CLASSIFICATION OF HYPOGLYCEMIC DISORDERS

A clinical classification of the more common causes of symptomatic hypoglycemia in adults is presented in Table 18–3. This classification is useful in directing diagnostic considerations. Symptomatic fasting hypoglycemia is a serious and potentially life-threatening problem warranting thorough evaluation. Conditions that produce inappropriate fasting hyperinsulinism are the most common cause of fasting hypoglycemia in otherwise healthy adults. These include insulin-secreting pancreatic β cell tumors and iatrogenic or surreptitious administration of insulin or sulfonylureas. In patients with illnesses that produce symptomatic fasting hypoglycemia despite appropriately suppressed insulin levels, the clinical picture is generally dominated by the signs and symptoms of the primary disease, with hypoglycemia often only a late or associated manifestation. Hypoglycemia, sometimes severe, can also occur postprandially. In patients with gastric surgery,

TABLE 18–3 Causes of symptomatic hypoglycemia in adults.

Fasting
With hyperinsulinism
Insulins
Oral insulin secretagogues (sulfonylureas, repaglinide, nateglinide)
Other drugs (pentamidine)
Autoimmune hypoglycemia (idiopathic insulin antibodies, insulin receptor autoantibodies)
Pancreatic β cell tumors
Without hyperinsulinism
Severe hepatic dysfunction
Chronic renal insufficiency
Inanition
Hypocortisolism
Alcohol use
Nonpancreatic tumors
Nonfasting
After gastric surgery
Functional
Noninsulinoma pancreatogenous hypoglycemic syndrome (NIPHS)
Occult diabetes
Ethanol ingestion with sugar mixers

overrapid gastric emptying and accelerated glucose absorption may result in excess insulin secretion, rapid disposal of glucose, and hypoglycemia. An uncommon condition of islet hyperplasia or adult nesidioblastosis called noninsulinoma pancreatogenous hypoglycemia syndrome can result in hypoglycemia 4 to 6 hours after meals.

SPECIFIC HYPOGLYCEMIC DISORDERS

1. DIABETES (SEE ALSO CHAPTER 17)

Iatrogenic hypoglycemia is common in type 1 patients and also in insulin treated type 2 patients. Most type 1 patients aiming for HbA_{1c} levels below 7% have on average one to two symptomatic hypoglycemic episodes per week. Severe hypoglycemia is defined as an episode requiring assistance, and in one study, incidence rates were about 12 per 100 patient years for both type 1 and insulin-treated type 2 patients. Sulfonylureas, repaglinide, and nateglinide can also cause hypoglycemia. Increased risk factors include age (70 years and older), renal failure, hepatic failure, and use of the long-acting sulfonylureas. A number of other drug–drug interactions (clarithromycin, salicylates, sulfonamides) can also potentiate the hypoglycemic effects of sulfonylureas. The annual incidence of sulfonylurea-induced hypoglycemia is approximately 0.2 per 1000 patient years.

As β cell failure progresses (early in type 1 and late in type 2 diabetes), patients lose their glucagon response to hypoglycemia. This combination of insulin deficiency and impaired glucagon response makes it harder for patients to achieve HbA_{1c} levels below 7% without occasional hypoglycemia. These hypoglycemic episodes attenuate the sympathoadrenal response to hypoglycemia, with decreased epinephrine release from the adrenal and decreased sympathetic neural responses (hypoglycemic unawareness) and impaired hepatic glucose release. This combination of events is termed hypoglycemia-associated autonomic failure, and it can persist for more than 24 hours after a single episode of hypoglycemia or longer after repeated episodes of hypoglycemia, which in turn increases the risk for recurrent hypoglycemia. About 20% to 25% of type 1 patients have hypoglycemic unawareness. Other factors that increase the risk for hypoglycemia include poor self-management skills. Patients may take too much insulin for the carbohydrates or high glucose adjustment or take the wrong insulin or do not appropriately time insulin administration with food ingestion. They also may not adjust the insulin for acute exercise or take extra carbohydrates for unexpected exercise or reduce insulin doses for improved insulin sensitivity with exercise training. Alcohol can decrease endogenous glucose production and can cause hypoglycemia, especially if it is consumed on an empty stomach. Diabetes complications—gastroparesis, autonomic neuropathy, and renal failure also increase the risk for hypoglycemia.

There are other consequences of hypoglycemia apart from the autonomic and neurogenic symptoms of acute hypoglycemia. In severe cases, hypoglycemia can cause convulsions and coma. Permanent neurological damage is rare. Although cross-sectional studies and case reports have reported intellectual decline with recurrent hypoglycemia, longitudinal studies have not shown significant cognitive dysfunction in adults. In the Diabetes Control and Complications Trial and its follow-up EDIC study, there was no evidence for cognitive decline related to hypoglycemia in 18 years of follow-up. Young children, however, may be more vulnerable to the effects of hypoglycemia on the brain. Hypoglycemia via its autonomic stimulation and catecholamine release increases cardiac output. In patients with cardiac disease, it can also precipitate cardiac arrhythmias, angina, myocardial infarction, and congestive heart failure. Unexpected deaths in type 1 patients are most likely due to hypoglycemia—so-called "dead-in-bed" syndrome. In studies from the United Kingdom and Scandinavian countries, the frequency of such unexplained deaths is about 2 to 6 events per 100,000 patient-years. Hypoglycemia can also exert a psychological toll. Acute hypoglycemia induces mood changes including anxiety and depression. Nocturnal hypoglycemia can lead to fatigue and decreased sense of well-being the following day. Patients who have had severe hypoglycemia may develop a phobia about hypoglycemia and keep their sugars unreasonably high. Hypoglycemia can also impact personal relationships, occupation, driving, and leisure activities. Surveys show that type 1 patients have increased risk of driving mishaps (crashes, moving violations) when compared to nondiabetic spouses; and that these are related to hypoglycemia.

The goal of therapy for hypoglycemia is to restore levels of plasma glucose to normal as rapidly as possible. If the patient is conscious and able to swallow, glucose-containing tablets or gels, or foods such as candy, juices (ie, orange or apple), and cookies should be quickly ingested. Fructose, found in many nutrient low-calorie sweeteners for diabetics, should not be used. While it can be metabolized by neurons, fructose is not transported across the blood–brain barrier.

If the patient is unconscious, rapid restoration of plasma glucose must be accomplished by giving 20 to 50 mL of 50% dextrose intravenously over 1 to 3 minutes (the treatment of choice) or, when intravenous glucose is not available, 1 mg of glucagon intramuscularly or intravenously. Families or friends of insulin-treated diabetics should be instructed in the administration of glucagon intramuscularly for emergency treatment at home. Glucagon should not be given if the hypoglycemia is due to sulfonylurea use. Under these circumstances, glucagon can stimulate insulin secretion and worsen the hypoglycemia. Attempts to feed the patient or to apply glucose-containing gels to the oral mucosa should be avoided because of the danger of aspiration. When consciousness is restored, oral feedings should be started immediately. In patients who have taken massive overdoses of sulfonylureas, the response to intravenous dextrose may be inadequate. For these patients, intravenous boluses of diazoxide (150-300 mg) may be tried but can result in hypotension. Intravenous octreotide (100 μg) has also been reported to be of benefit.

Patients on insulin or sulfonylureas should be instructed on how to recognize and treat hypoglycemia and what measures they can take to prevent such episodes. Patients with type 1 and insulin-treated type 2 diabetes should monitor their blood glucose frequently. Hypoglycemia not infrequently occurs at night and patients should avoid taking large doses of short-acting insulin just

before going to bed. Patients should from time to time also monitor blood glucose levels in the middle of the night. Hypoglycemia can also occur many hours after strenuous exercise, and patients should be advised to monitor at these times and cut back their insulin doses and/or eat more carbohydrate. Continuous glucose monitoring systems are increasingly used by patients to alert them to falling glucose levels and prevent hypoglycemia. Finally, it is important to individualize glycemic goals. Early in the course of both type 1 and type 2 diabetes when there is still some endogenous β cell function, it is easier to achieve HbA_{1c} levels close to normal with low risk of hypoglycemia. As β cell failure progresses, however, aiming for normality may lead to unacceptably high rates of hypoglycemia. Patients who have had frequent hypoglycemia and have hypoglycemia unawareness should be encouraged to temporarily raise their glycemic targets—as little as 2 to 3 weeks of scrupulous avoidance of hypoglycemia reverses hypoglycemia unawareness and improves the attenuated epinephrine response. Diabetes complications, previous incidence of hypoglycemia, and life expectancy should all be considered in the establishment of glycemic goals.

2. FACTITIOUS HYPOGLYCEMIA

Factitious hypoglycemia should be suspected in any patient with access to insulin or sulfonylurea drugs. It is most commonly seen in health professionals and patients with diabetes or their relatives. The reasons for self-induced hypoglycemia vary, with many patients having severe psychiatric disturbances or a pathological need for attention. Inadvertent ingestion of sulfonylureas resulting in clinical hypoglycemia has also been reported, due either to patient error or to a prescription mishap on the part of a pharmacy.

When insulin is used to induce hypoglycemia, an elevated serum insulin level often raises suspicion of an insulin-producing pancreatic β cell tumor. It can be difficult to prove that the insulin is of exogenous origin. The combination of hypoglycemia, high immunoreactive insulin levels, suppressed plasma C-peptide, and suppressed proinsulin level is pathognomonic of exogenous insulin administration in nondiabetic patients. Insulin and C-peptide are secreted into the portal circulation in a 1:1 molar ratio. The liver is the primary site of insulin clearance, and a large fraction of the portal insulin is removed by the liver during first pass transit. C-peptide, on the other hand, is mostly cleared by the kidney at a slower rate than insulin. It follows that normally the molar ratio of insulin to C-peptide is less than 1.0. In the case of factitious hyperinsulinemia, the insulin/C-peptide ratio (pmol/L) will be greater than 1.0. Patients with renal failure may have normal or even high plasma C-peptide levels, but plasma proinsulin levels are suppressed.

When sulfonylurea abuse is suspected, plasma or urine should be screened for its presence. Hypoglycemia with inappropriately elevated levels of serum insulin *and* C-peptide along with detectable sulfonylureas in blood or urine is diagnostic of inadvertent or factitious sulfonylurea overdose. It is important to use assays which measure not only all the sulfonylureas but also repaglinide and nateglinide.

Patients with factitious hypoglycemia should receive psychiatric treatment and/or psychotherapy.

3. DRUGS

Numerous pharmacologic agents may potentiate the effects of insulin and predispose to hypoglycemia. Common offenders include fluoroquinolones such as gatifloxacin and levofloxacin, pentamidine, quinine, angiotensin-converting enzyme (ACE) inhibitors, ethanol, salicylates, and beta-adrenergic–blocking drugs. The fluoroquinolones, especially gatifloxacin, act on ATP-sensitive potassium channels (K_{ATP}) in the β cell and initially cause hypoglycemia and then hyperglycemia after several days into therapy. Intravenous pentamidine is cytotoxic to pancreatic β cells, resulting in acute insulin release and hypoglycemia. This occurs in about 10% to 20% of patients receiving the drug and may be followed later by persistent insulinopenia and hyperglycemia. Fasting patients taking noncardioselective beta blockade can have an exaggerated hypoglycemic response to starvation. Beta blockade inhibits fatty acid and gluconeogenic substrate release and reduces plasma glucagon levels resulting in hypoglycemia. Also the symptomatic response to hypoglycemia is altered—tachycardia is blocked while hazardous elevations of blood pressure may result during hypoglycemia in response to the unopposed alpha-adrenergic stimulation from circulating catecholamines and neurogenic sympathetic discharge. Symptoms of sweating, hunger, and uneasiness are not masked by beta-blocking drugs and remain indicators of hypoglycemia in the aware patient.

Therapy with ACE inhibitors increases the risk of hypoglycemia in diabetic patients who are taking insulin or sulfonylureas, presumably because these drugs increase sensitivity to circulating insulin by increasing blood flow to muscle.

Ethanol-associated hypoglycemia has been proposed to occur as a consequence of hepatic alcohol dehydrogenase activity depleting NAD. The resultant change in the redox state (increase in NADH-to-NAD$^+$ ratio) causes a partial block at several points in the gluconeogenic pathway. These include depressed conversion of citric acid cycle metabolites to oxaloacetate; decreased conversion of lactate to pyruvate; and decreased conversion of L-alpha-glycerophosphate to dihydroxyacetone-phosphate.

In the patient who is imbibing ethanol but not eating, fasting hypoglycemia may occur after hepatic glycogen stores have been depleted (within 8-12 hours of a fast). No correlation exists between the blood ethanol levels and the degree of hypoglycemia, which may occur while blood ethanol levels are declining. It should be noted that ethanol-induced fasting hypoglycemia may occur at ethanol levels as low as 45 mg/dL (10 mmol/L)—considerably below most states' legal standards (80 mg/dL [17.4 mmol/L]) for being *under the influence*. Most patients present with neuroglycopenic symptoms, which may be difficult to differentiate from the neurotoxic effects of the alcohol. These symptoms in a patient whose breath smells of alcohol may be mistaken for alcoholic stupor. Intravenous dextrose should be administered promptly to all such stuporous or comatose patients. Because hepatic glycogen stores have been depleted by the time hypoglycemia occurs, parenteral glucagon is not

effective. Adequate food intake during alcohol ingestion prevents this type of hypoglycemia.

4. AUTOIMMUNE HYPOGLYCEMIA

In recent years, a rare autoimmune disorder has been reported in which patients have circulating insulin antibodies and the paradoxic feature of hypoglycemia. More than 200 cases of insulin-antibody-associated hypoglycemia have been reported since 1970, with 90% of cases reported in Japanese patients. HLA class II alleles (DRB1*0406, DQA1*0301, and DQB1*0302) are associated with this syndrome, and these alleles are 10 to 30 times more prevalent in Japanese and Koreans, which may explain the higher prevalence of this syndrome in these populations. Hypoglycemia generally occurs 3 to 4 hours after a meal and follows an early postprandial hyperglycemia. It is attributed to a dissociation of insulin-antibody immune complexes, releasing free insulin. This autoimmune hypoglycemia, which is due to accumulation of high titers of antibodies capable of reacting with endogenous insulin, has been most commonly reported in methimazole-treated patients with Graves disease from Japan as well as in patients with various other sulfhydryl-containing medications (captopril, penicillamine) and other drugs such as hydralazine, isoniazid, procainamide, and alpha lipoic acid. In addition, it has been reported in patients with autoimmune disorders such as rheumatoid arthritis, systemic lupus erythematosus, and polymyositis, as well as in multiple myeloma and other plasma cell dyscrasias where paraproteins or antibodies cross-react with insulin.

High titers of insulin autoantibodies, usually IgG class, can be detected. Insulin, proinsulin, and C-peptide levels may be elevated, but the results may be erroneous because of the interference of the insulin antibodies with the immunoassays for these peptides.

In most cases the hypoglycemia is transient and usually resolves spontaneously within 3 to 6 months of diagnosis, particularly when the offending medications are stopped. The most consistent therapeutic benefit in management of this syndrome has been achieved by dietary treatment with frequent, low-carbohydrate, small meals. Prednisone therapy (30-60 mg/d) has been used to lower the titer of insulin antibodies.

Hypoglycemia due to insulin receptor autoantibodies is also an extremely rare syndrome; most cases have occurred in women, often with a history of autoimmune disease. Almost all of these patients have also had episodes of insulin-resistant diabetes and acanthosis nigricans. Their hypoglycemia may be either fasting or postprandial and is often severe and is attributed to an agonistic action of the antibody on the insulin receptor. Balance between the antagonistic and agonistic effects of the antibodies determines whether insulin-resistant diabetes or hypoglycemia occurs. Hypoglycemia was found to respond to glucocorticoid therapy but not to plasmapheresis or immunosuppression.

5. PANCREATIC β CELL TUMORS

Spontaneous fasting hypoglycemia in an otherwise healthy adult is most commonly due to an insulinoma, an insulin-secreting tumor of the islets of Langerhans. Eighty percent of these tumors are single and benign; 10% are malignant (if metastases are identified); and the remainder are multiple, with scattered micro- or macroadenomas interspersed within normal islet tissue. (As with some other endocrine tumors, histologic differentiation between benign and malignant cells is difficult, and close follow-up is necessary to ensure the absence of metastases.)

These adenomas may be familial and have been found in conjunction with tumors of the parathyroid glands and the pituitary (multiple endocrine neoplasia type 1; see Chapter 22). It has been reported that about 30% of sporadic insulinoma tumors have a somatic mutation in the YY1 gene (T372R) that encodes the transcriptional repressor YY1. Over 99% of them are located within the pancreas and less than 1% in ectopic pancreatic tissue.

These tumors may appear at any age, although they are most common in the fourth to sixth decades. A slight predominance in women has been reported in some studies. However, others suggest that there is no gender predilection.

Clinical Findings

Symptoms and signs   The most important prerequisite to diagnosing an insulinoma is simply to consider it, particularly when facing a clinical presentation of hypoglycemia with symptoms of central nervous system dysfunction such as confusion or abnormal behavior. Whipple's triad consists of history of symptoms consistent with hypoglycemia, associated low plasma glucose and relief of symptoms upon ingesting carbohydrates and raising the plasma glucose.

Patients typically complain of feeling tired, experiencing blurred vision, and not thinking clearly. Other symptoms include personality changes, amnesia, and loss of consciousness. The preponderance of neuroglycopenic symptoms, rather than those more commonly associated with the adrenergic symptoms of hypoglycemia can lead to misdiagnoses of psychiatric or neurological disorders. Some patients learn to relieve or prevent their symptoms by taking frequent feedings. Eating or drinking readily absorbable carbohydrates improves the symptoms within approximately 15 minutes. Patients may gain weight but obesity is seen in less than 30% of patients with insulin-secreting tumors.

Timing of the symptoms in relation to meals and exercise should be noted. In insulinoma patients, the symptoms are most likely to occur early in the morning before breakfast or if a meal is missed during the day. Patients occasionally present when they attempt to go on a diet to lose weight. Exercise may precipitate the symptoms, especially if the activity occurs while fasting or some hours after a meal. Some patients develop symptoms in the middle of the night and may have to eat. Partners of patients often provide useful information, especially if the patient requires assistance in treating the hypoglycemic symptoms. Occasionally the emergency services are called when the patient gets severely confused, presents with focal weakness, loses consciousness, or has a seizure. There may be documentation of low fingerstick glucose at the time when symptoms are present with recovery following administration of intravenous glucose. With the ready availability of home blood glucose–monitoring systems, patients sometimes present with documented fingerstick blood glucose levels in 40s

and 50s at time of symptoms. Access to diabetic medications (sulfonylureas or insulin) should be explored—does a family member have diabetes, or does the patient or family member work in the medical field? Medication-dispensing errors should be excluded—has the patient's prescription medication changed in shape or color? Other illnesses that cause hypoglycemia such as renal failure, hepatic failure, Addison disease, or nonislet tumor should be considered. Patients with insulinoma or factitious hypoglycemia usually have a normal physical examination.

Diagnostic Testing

If the history is consistent with episodic spontaneous hypoglycemia, patients should be given a home blood glucose monitor and advised to monitor blood glucose levels at the time of symptoms, if it is safe, and before consumption of carbohydrates. Patients with insulinomas frequently will report fingerstick blood glucose levels between 40 mg/dL (2.2 mmol/L) and 50 mg/dL (2.8 mmol/L) at the time of symptoms. The diagnosis, however, cannot be made based on a fingerstick blood glucose since they are not sufficiently accurate—the fingerstick value may differ by as much as 15 mg/dL when the true blood glucose is less than 75 mg/dL. It is necessary to have a low laboratory glucose concomitantly with elevated plasma insulin, proinsulin and C-peptide levels, and a negative sulfonylurea screen. When patients give a history of symptoms after only a short period of food withdrawal or with exercise, then an outpatient assessment can be attempted. The patient should be brought by a family member to the office after an overnight fast and observed in the office. Activity such as walking should be encouraged and fingerstick blood glucose measured repeatedly during observation. If symptoms occur or fingerstick blood glucose is below 50 mg/dL then samples for plasma glucose, insulin, C-peptide, proinsulin, sulfonylurea screen, serum ketones, and antibodies to insulin should be sent. If outpatient observation does not result in symptoms or hypoglycemia and if the clinical suspicion remains high, then the patient should undergo an inpatient, supervised, 72-hour fast.

A suggested protocol for the supervised fast is set forth in Table 18–4. The fast is timed from the last intake of calories. The term *72-hour fast* is actually a misnomer in most cases, because the fast should be immediately terminated as soon as symptoms and laboratory confirmation of hypoglycemia are evident. About 43% of patients with insulinomas are symptomatic within the first 12 hours, 67% by 24 hours and 93% to 95% by 48 hours. Since a minority of patients (~5%-7%) may not demonstrate hypoglycemia after 48 hours of fasting, it is preferable to continue the fast for up to 72 hours. Patients can drink water and decaffeinated, noncaloric drinks. It is important that patients are active (walking) during the fast since exercise may help precipitate hypoglycemia. An intravenous cannula should be placed to allow for blood draws and as a safety precaution, should intravenous dextrose infusion be required. Fingerstick blood glucose levels should be measured at intervals, and blood sent to the laboratory, if the patient is symptomatic or when fingerstick blood glucose levels are below 50 mg/dL. The decision to stop the fast is not always easy. Patients should be carefully monitored for symptoms of hypoglycemia. If the patient

TABLE 18–4 Suggested hospital protocol for supervised rapid diagnosis of insulinoma.

(1) Obtain baseline plasma glucose, insulin, proinsulin, betahydroxybutyrate and C-peptide measurements at onset of fast and place intravenous cannula.

(2) Permit only calorie-free and caffeine-free fluids and encourage supervised activity (such as walking).

(3) Obtain capillary glucose measurements with a reflectance meter every 4 hours until values <60 mg/dL are obtained. Then increase the frequency of fingersticks to each hour, and when capillary glucose value is <45 mg/dL, send a venous blood sample to the laboratory for serum glucose, insulin, proinsulin, betahydroxybutyrate, and C-peptide measurements. Check frequently for manifestations of neuroglycopenia.

(4) If symptoms of hypoglycemia occur *or* if a laboratory value of plasma glucose is <45 mg/dL *or* if 72 hours have elapsed, then conclude the fast with a final blood sample for serum glucose, insulin, proinsulin, C-peptide, β-hydroxybutyrate, and sulfonylurea measurements. Then give oral fast-acting carbohydrate followed by a meal. If the patient is confused or unable to take oral agents, administer 50 mL of 50% dextrose intravenously over 3-5 minutes. Do not conclude a fast based simply on basis of a capillary blood glucose measurement—wait for the laboratory glucose value unless the patient is symptomatic and it would be dangerous to wait.

Glucose sample should be collected in sodium fluoride containing tube on ice to prevent glycolysis and the plasma separated immediately upon receipt at the laboratory. Arrange for the laboratory to run the glucose samples "stat."

is symptomatic and the laboratory glucose is less than 45 mg/dL, then the test can be stopped. If the symptoms are equivocal and the laboratory glucose is in the mid 50s or higher, then the fast should be continued, if the patient is agreeable. At the time of termination of the fast, blood should be sent to the laboratory for plasma glucose, insulin, proinsulin, C-peptide, and serum β-hydoxybutyrate levels and a sulfonylurea screen.

In normal men, the blood glucose value does not fall below 55 mg/dL (3.1 mmol/L) during a 72-hour fast, whereas insulin levels fall below 10 μU/mL; in some normal women, however, plasma glucose may fall below 30 mg/dL (1.7 mmol/L) (lower limits have not been established), but serum insulin levels also fall appropriately to less than 5 μU/mL. These women remain asymptomatic despite this degree of hypoglycemia, presumably because ketogenesis is able to provide sufficient fuel for maintenance of central nervous system function.

The diagnostic criteria for insulinoma after a 72-hour fast are listed in Table 18–5. Virtually all patients with insulin-secreting islet cell tumors fail to suppress their insulin secretion appropriately when the plasma glucose is less than 45 mg/dL (see Table 18–5). It is important to be aware of limitations of the particular insulin assay that is used. When radioimmunoassays (RIAs) with a sensitivity of 5 mU/mL are used, patients with insulinomas have plasma insulin concentrations of 6 mU/mL or more. Immunochemiluminometric assays (ICMA) have sensitivities of less than 1 mU/mL, and with these assays, the cutoff for insulinomas is 3 mU/mL or higher. Previous treatment with insulin may lead to development of autoantibodies to insulin that interfere with the

TABLE 18–5 Diagnostic criteria for insulinoma after a 72-hour fast.

Plasma glucose	<45 mg/dL (2.5 mmol/L)
Plasma insulin (RIA)	≥6 uU/mL (36 pmol/L)
Plasma insulin (ICMA)	≥3 uU/mL (18 pmol/L
Plasma C-peptide	≥200 pmol/L (0.2 nmol/L; 0.6 ng/mL)
Plasma proinsulin	≥5 pmol/L
β-Hydroxybutyrate	≤2.7 nmol/L
Sulfonylurea screen (including repaglinide and nateglinide)	Negative

Conversion factors: insulin uU/mL × 6.0 = pmol/L; C-peptide ng/mL × 0.33 = nmol/L.

Abbreviations: RIA, radioimmunoassay; ICMA, Immunochemiluminometric assays.

assays, leading to falsely low or elevated values depending on the method used. Proper collection of samples is also important. If serum is not separated and frozen within 1 to 2 hours, falsely low values result because the insulin molecule undergoes proteolytic degradation. Plasma insulin levels measured by ICMA may be lower in hemolyzed samples. Calculation of ratios of insulin (μU/mL) to plasma glucose (mg/dL) is not helpful in making the diagnosis.

Factitious use of insulin will result in suppression of endogenous insulin secretion and low C-peptide levels. Elevated circulating proinsulin levels in the presence of fasting hypoglycemia is characteristic of β cell tumors and does not occur in factitious hyperinsulinism. Thus, C-peptide and proinsulin levels (by ICMA) of greater than or equal to 200 pmol/L and greater than or equal to 5 pmol/L, respectively, are characteristic of insulinomas. C-peptide is renally cleared and caution should be used in interpreting elevated levels in the setting of renal failure. Proinsulin normally represents less than 10% of total immunoreactive insulin. Insulinoma cells are poorly differentiated, which affects their ability to process proinsulin to insulin. Thus, most patients with insulinoma have elevated levels of proinsulin, representing as much as 30% to 90% of total immunoreactive insulin. Hyperinsulinemia suppresses ketone production. Plasma β-hydroxybutyrate levels in patients with insulinoma are 2.7 nmol/L or less. A progressive increase in β-hydroxybutyrate levels after 18 hours of fasting is strongly predictive that the fast will be negative. No single hormone measurement (insulin, proinsulin, C-peptide) is 100% sensitive and specific for the diagnosis of insulinoma, and insulinoma cases have been reported with insulin levels below 3 μU/mL (ICMA assay) or proinsulin level below 5 pmol/L. The hormonal assays are also not standardized between labs, and there can be significant variation in the results. Therefore, the diagnosis should be based on multiple biochemical parameters.

A variety of stimulation tests with intravenous tolbutamide, glucagon, or calcium have been devised to demonstrate exaggerated and prolonged insulin secretion in the presence of insulinomas. However, because insulin-secreting tumors have a wide range of granule content and degrees of differentiation, they are variably responsive to these secretagogues. Thus, absence of an excessive insulin-secretory response during any of these stimulation tests does not rule out the presence of an insulinoma. In addition, the tolbutamide stimulation test is extremely hazardous to patients with responsive tumors because it induces prolonged and refractory hypoglycemia. For that reason, it is no longer included in the diagnostic workup of insulinoma.

The hyperinsulinism associated with insulinoma impairs glycogenolysis. Thus, when patients with insulinoma are given 1 mg IV glucagon at the end of the 72-hour fast, there is an increase in plasma glucose within the first 30 minutes. The glucose rise is 25 mg/dL (1.4 mmol/L) or more at 20 to 30 minutes, whereas normal subjects have a lower increment. Intravenous glucagon can also cause an exaggerated release of insulin from insulinomas. Patients are tested after an overnight fast, and serum insulin levels measured every 5 minutes for 15 minutes after 1 mg IV glucagon. A stimulated peak insulin level exceeding 130 μU/mL (780 pmol/L) (~twice upper limit of peak stimulated normals)\suggests an insulin-secreting tumor. However, only about half of patients with insulinomas have insulin levels above 130 μU/mL and so the test is not so helpful. Also, in some patients the exaggerated insulin secretion can lead to severe hypoglycemia. Nausea is an unpleasant side effect, often occurring several minutes after administration of intravenous glucagon.

The oral glucose tolerance test is of no value in the diagnosis of insulin-secreting tumors. A common misconception is that patients with insulinomas have flat glucose tolerance curves because the tumors discharge insulin in response to oral glucose. In fact, most insulinomas respond poorly, and curves typical of diabetes are more common. In those rare tumors that do release insulin in response to glucose, a flat curve may result; however, this also can be seen occasionally in normal subjects.

Low HbA$_{1c}$ values have been reported in patients with insulinoma, reflecting the presence of chronic hypoglycemia. There is however considerable overlap with normal patients and no HbA$_{1c}$ value is diagnostic.

Tumor Localization Studies

After the diagnosis of insulinoma has been unequivocally made by clinical and laboratory findings, studies to localize the tumor should be initiated. The focus of attention should be directed at the pancreas only, because virtually all insulinomas originate from this tissue. Ectopic cases are very rare. There are reports of duodenal neuroendocrine tumors secreting insulin. Other ectopic sites include spleen, perisplenic tissue, duodenohepatic ligament, and ligament of Treitz. Ovarian carcinomas and teratomas, small cell carcinomas of the cervix, and bronchial carcinoids have also been reported to secrete insulin. Because of the small size of these tumors (average diameter of 1.5 cm in one large series), imaging studies do not necessarily identify all tumors. A pancreatic dual phase, thin section, helical CT scan can identify 82% to 94% of the lesions. MRI scans with gadolinium can be helpful in detecting a tumor in 85% of cases. One case report suggests that diffusion-weighted MRI can be useful for detecting and localizing small insulinomas, especially those with no hypervascular pattern. ^{111}In-octreotide scans for insulinomas, which typically express somatostatin receptor type 3, are only positive in 50% to 70% of cases. Newer PET/CT scans, using gallium-labeled somtatostatin analogs such as DOTA-1-NaI[3]-octreotide (DOTA-NOC), which
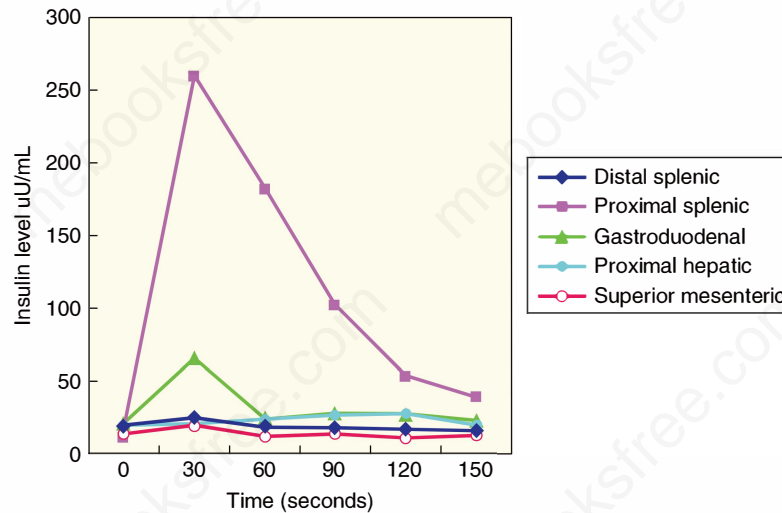
FIGURE 18–4 Responses of serum insulin to selective intra-arterial calcium stimulation in a patient with biochemical confirmation of inappropriate hyperinsulinism. 0.0125 mmol/kg calcium gluconate diluted in 5 mL normal saline were bolused into selectively catheterized arteries as indicated and blood samples collected from the hepatic vein at 0, 30, 50, 90, 120, and 180 seconds. An islet cell tumor was removed from the tail of the pancreas.

has a higher affinity for somatostatin receptor subtypes 2, 3, and 5, have been reported to be useful in localizing the tumors. Insulinomas express GLP1-receptors and radiolabeled GLP1-receptor agonists such as [Lys(40)(Ahx-hydrazinonicotinamide [HYNIC]-(99m)Tc)NH(2)]-exendin-4 for SPECT/CT have also been reported to visualize the tumors. The optimal imaging study used will depend upon local availability and local radiologic skill. If the imaging study is negative, then an endoscopic ultrasound should be performed. In experienced hands, about 80% to 90% of tumors can be detected. Finally, needle aspiration of the identified lesion can be attempted to confirm the presence of a neuroendocrine tumor. If the tumor is not identified or imaging results are equivocal, then the patient should undergo selective calcium-stimulated angiography, which has been reported to localize the tumor to particular regions of the pancreas approximately 90% of the time. In this test, angiography is combined with injections of calcium gluconate into the gastroduodenal, splenic, and superior mesenteric arteries, and insulin levels are measured in the hepatic vein effluent. The procedure is performed after an overnight fast. Ten percent of calcium gluconate, diluted to a volume of 5 mL normal saline, is injected into the individual arteries (discussed earlier) at a dose of 0.0125 mmol Ca^{2+}/kg body weight (0.005 mmol/kg for obese patients). Five milliliter blood samples are taken from the hepatic effluent at times 0, 30, 60, 90, 120, and 180 seconds after calcium injection. Fingerstick blood glucose levels are measured at intervals and a dextrose infusion is maintained throughout the procedure. Calcium stimulates insulin release from insulinomas but not normal islets. A step-up in insulin levels at 30 or 60 seconds (twofold or greater) regionalizes the source of the hyperinsulinism to the head of the pancreas for the gastroduodenal artery, the uncinate process for the superior mesenteric artery, and the body and tail of the pancreas for the splenic artery. A less than twofold elevation of insulin in the 120 second sample may represent effects of recirculating calcium and is not considered a

positive localization. In a single insulinoma, the response is in one artery alone (Figure 18–4) unless the tumor resides in an area fed by two arteries or if there are multiple insulinomas as in multiple endocrine neoplasia, type 1. Patients who have diffuse islet hyperplasia (the noninsulinoma pancreatogenous hypoglycemia syndrome [NIPHS]) will have positive responses in multiple arteries. Because diazoxide may interfere with this test, it should be discontinued for at least 48 to 72 hours before sampling. Patients should be closely monitored during the procedure to avoid hypoglycemia (as well as hyperglycemia), which could affect insulin gradients. These studies combined with careful intraoperative ultrasonography and palpation by a surgeon experienced in insulinoma surgery correctly identify up to 98% of tumors.

Treatment of Insulinoma

The treatment of choice for insulin-secreting tumors is surgical resection (Figure 18–5). While waiting for surgery, patients should be given diazoxide, a potent inhibitor of insulin secretion. It acts by opening the K_{ATP} channel of the pancreatic β cell and hyperpolarizing the cell membrane. This reduces calcium influx through the voltage-gated calcium channel, thereby reducing insulin release. Divided doses of 300 to 400 mg/d usually suffice, but occasionally a patient may require up to 800 mg/d. A liquid preparation of diazoxide (50 mg/mL) is available in the United States. Side effects include edema due to sodium retention (which generally necessitates concomitant thiazide administration), gastric irritation, and mild hirsutism.

Tumor resection should be performed only by surgeons with extensive experience with removal of islet cell tumors, because these tumors may be small and difficult to recognize. The tumors are enucleated whenever possible unless they have malignant features (eg, hardness or an appearance of infiltration). Preoperative imaging studies, including endoscopic ultrasound, can identify
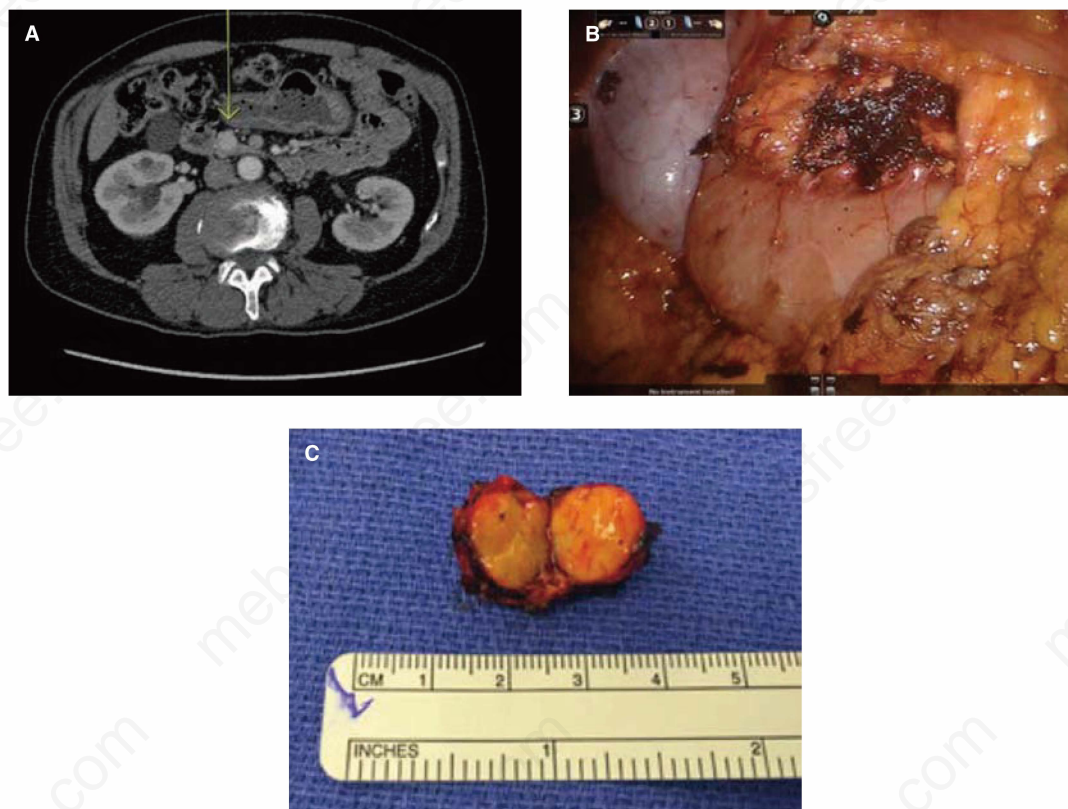
FIGURE 18–5 1.4 cm insulinoma resected from the uncinate process of the pancreas. **A:** CT scan of abdomen; **B:** View of the 3rd part of the duodenum (bottom) and uncinate process with site of enucleation; **C:** tumor sectioned in half.

tumors amenable to laparoscopic surgery. Laparoscopic surgery is associated with faster postoperative recovery. Laparoscopic intra-operative ultrasound should be used to confirm the location and depth of the tumor within the pancreas and also note its relation-ship to the pancreatic duct and splenic vessels. Open surgery is still necessary for some tumors such as those in the head of the pan-creas close to the main pancreatic duct. In the very occasional case where the tumor cannot be found at operation despite the use of intraoperative ultrasound, it is no longer advisable to blindly resect the body and tail of the pancreas since a nonpalpable tumor, missed by ultrasound, is most likely embedded within the fleshy head of the pancreas that is left behind with subtotal resections. Most surgeons prefer to close the incision and treat the patient medically and/or repeat the localization studies. Ultrasound-guided ethanol injection into the tumor has been reported to be effective and should be considered for those patients who are poor surgical candidates.

Diazoxide should be administered on the day of surgery in patients who are responsive to it, because the drug greatly reduces the need for glucose supplements and the risk of hypoglycemia during surgery. Typically, it does not mask the glycemic rise indicative of surgical cure. Blood glucose levels should be moni-tored frequently during the operation, and a 5% or 10% dextrose infusion should be used to maintain euglycemia. Hyperglycemia occurs for a few days postoperatively most likely due to edema and inflammation of the pancreas secondary to its mobilization and

manipulation during surgical resection of the insulinoma. How-ever, other possible contributing factors include high levels of counterregulatory hormones induced by the procedure, chronic downregulation of insulin receptors by the previously high circu-lating insulin levels from the tumor, and, perhaps, suppression of normal pancreatic β cells by long-standing hypoglycemia. Small subcutaneous doses of regular insulin may be prescribed every 4 to 6 hours if plasma glucose exceeds 300 mg/dL (16.7 mmol/L), but in most cases pancreatic insulin secretion recovers after 48 to 72 hours, and very little insulin replacement is required.

In patients with inoperable functioning islet cell carcinoma with or without hepatic metastasis and in approximately 5% to 10% of MEN 1 cases when subtotal removal of the pancreas has failed to produce cure, the treatment approach is the same as for other types of pancreatic neuroendocrine tumors (PNET). Diazoxide is the treatment of choice in preventing hypoglycemia. Frequent carbohydrate feedings (every 2-3 hours) can also be helpful although weight gain can become a problem. If patients are unable to tolerate diazoxide because of gastrointestinal upset, hirsutism, or edema, the calcium channel-blocker verapamil, which inhibits insulin release from insulinoma cells, can be tried. Somatostatin analogs, such as octreotide or lanreotide, should be considered if diazoxide is ineffective or if there is tumor progres-sion. These drugs not only inhibit insulin secretion but also have antiproliferative activity. Surgery or embolization (bland-, chemo-, or radio-embolization) or thermal ablation (radiofrequency,

microwave, or cryoablation) can be used to reduce tumor burden and also provide symptomatic relief. For islet cell carcinomas, chemotherapy regimens can be considered. Representative regimens include combinations of streptozocin, 5–fluorouracil and doxorubicin; capecitabine and oxaliplatin; and capecitabine and temozolomide. There are only limited data regarding efficacy of these regimens. Targeted therapies against multiple steps in the PI3K/AKT/mTor pathway have been shown to be helpful. Everolimus, an inhibitor of mTor, has been approved for treatment of advanced pancreatic neuroendocrine tumors. Sunitinib targets multiple tyrosine kinase inhibitors including vascular endothelial growth factor (VEGF) receptors 2 and 3, platelet-derived growth factor (PDGF) receptors α and β, and c-kit, has been shown to slow growth of pancreatic neuroendocrine tumors. Treatment with radioisotopes (indium-111, yttrium-90, or lutetium-177) linked to a somatostatin analog have been reported to show benefit in a proportion of patients.

6. NONISLET CELL TUMOR HYPOGLYCEMIA (NICTH)

A variety of nonislet cell tumors have been found to cause fasting hypoglycemia. Most are large and mesenchymal in origin, retroperitoneal fibrosarcoma being the classic prototype. However, hepatocellular carcinomas, adrenocortical carcinomas, renal cell carcinomas, gastrointestinal tumors, lymphomas, leukemias, and a variety of other tumors have also been reported.

Laboratory diagnosis depends on fasting hypoglycemia associated with serum insulin levels below 5 μU/mL. In many cases, the hypoglycemia is due to the expression and release of an incompletely processed insulin-like growth factor II (IGF-II) by the tumor (see also Chapter 21). The primary IGF-II translation product is pre-pro-IGF-II consisting of N-terminal signal peptide of 24 amino acids, 67 amino acid mature IGF-II, and an 89 amino acid extension (E-domain) at the C-terminus. Posttranslational processing of pre-pro-IGF-II involves removal of the N-signal sequence, O-glycosylation of one or more threonine residues of the E-domain, and sequential proteolysis of the E-domain. During this process an IGF-II protein with a 21 amino extension of the E-domain (pro-IGF-IIE[68-88]) is a relatively stable intermediate that may be secreted from the cell. Most of the mature IGF-II released from the liver is complexed with IGF-binding protein-3 (IGFBP3) and acid-labile subunit (ALS). This ternary protein complex is generally inactive in adults because it is unable to bind properly to tissue receptors. It is only the free IGF-II (<1%) and that bound in binary complexes (predominantly IGFBP2 and IGFBP3) that is accessible to tissue compartment and available to bind the IGF and insulin receptors. However, in patients with nonpancreatic tumors associated with hypoglycemia, incompletely processed mainly nonglycosylated forms of IGF-II are released—in particular pro-IGF-IIE (68-88) form. These incompletely processed molecules are heterogeneous in size and are also referred to as **big IGF-II** and have molecular mass of 10 to 17 kDa in contrast to mature IGF-II at 7.5 kDa. Pro-IGF-II can form binary complexes with IGFBPs but have reduced affinity

for forming a tertiary complex with ALS. As a consequence, more of the pro-IGF-II is available for binding to the insulin receptors in the muscle to promote glucose transport and to insulin receptors in liver and kidney to reduce glucose output. The increased production of pro-IGF-II by the tumor may also displace processed IGF-II from IGFBPs, increasing free, unbound IGF-II. The IGF-II may bind to receptors for IGF-I in the pancreatic β cell to inhibit insulin secretion and in the pituitary to suppress growth hormone release. With the reduction of growth hormone, there is a consequent lowering of IGF-I levels as well as IGFBP-3 and ALS.

Size exclusion acid chromatography has been the standard method for detection of Pro-IGF-II in NICTH, but the process is time-consuming. Immunoblot analysis after separating the proteins on 16.5% tricine-sodium dodecyl sulphate-polyacrylamide gels is a more rapid and equally sensitive method. The IGF-II antibody used recognizes both mature and pro-IGF-II forms. In normal subjects, most of the IGF-II migrates at 7.5 kDa and a small amount in the 10 to 17 kDa region, whereas with NICTH most of the IGF-II migrates in the 10 to 17 kDa region and a small amount at 7 kDa.

The clinical syndrome of nonislet cell tumor hypoglycemia, therefore, is supported by laboratory documentation of serum insulin levels below 5 μU/mL with plasma glucose measurements of 45 mg/dL or lower. Values for growth hormone and IGF-I are also decreased. Levels of IGF-II may be increased but often are normal in quantity despite the presence of the immature, higher molecular weight form of IGF-II, which can only be detected by special laboratory techniques.

Not all the patients with NICTH have elevated pro-IGF-II. Ectopic insulin production has been described (bronchial carcinoid, ovarian carcinoma, teratoma, and small cell carcinoma of the cervix). Hypoglycemia due to IGF-I released from a metastatic large cell carcinoma of the lung has also been reported.

Treatment is aimed toward the primary tumor, with supportive therapy using frequent feedings. Diazoxide is ineffective in reversing the hypoglycemia caused by these tumors.

7. POSTPRANDIAL HYPOGLYCEMIA

Hypoglycemia Following Gastric Surgery

Hypoglycemia sometimes occurs after gastric surgery (gastrectomy, vagotomy, pyloroplasty, gastrojejunostomy, and laparoscopic Nissen fundoplication, Billroth II procedure, and Roux-en-Y gastric bypass), often after patients consume foods containing high levels of carbohydrates. This late dumping syndrome occurs 1 to 3 hours after a meal and is result of rapid delivery of high concentration of carbohydrates in the small bowel and rapid increase in blood glucose. This is countered by a hyperinsulinemic response. The high insulin levels are responsible for the subsequent hypoglycemia. The hypoglycemic symptoms include lightheadedness, sweating, confusion, and even loss of consciousness. It is likely that excessive release of gastrointestinal hormones—such as glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide 1 (GLP-1)—may play a role in

the hyperinsulinemic response. An increased GLP-1 response has been noted in patients after total gastrectomy, esophageal resection, partial gastrectomy, and Roux-en-Y surgery; and a positive correlation has been noted between the rise in plasma GLP-1 and insulin release. It has been reported that treatment with exendin 9-39, a GLP-1 receptor antagonist can prevent post-gastric bypass hypoglycemia. Decreased need for gastric surgery for peptic ulcer disease and newer gastric operations such as proximal gastric vagotomy had reduced the incidence of postgastrectomy syndromes. There has been a resurgence of cases, however, with the popularity of Roux-en-Y gastric bypass surgery for morbid obesity. The prevalence of the syndrome after Roux-en-Y procedure is not known. The University of Minnesota surgery group identified 14 cases of postprandial hypoglycemia in 3082 procedures (0.4%), but it is unclear how many patients may have eluded detection or failed to maintain follow-up.

Patients typically complain of more severe symptoms after consumption of large amounts of readily absorbable carbohydrates. The mixed meal test can be used to precipitate the symptoms. The composition of the mixed meal has not been standardized, and it is reasonable to request the patient to consume a meal that leads to symptoms during everyday life. The University of Minnesota investigators formalized a high carbohydrate and a low carbohydrate test meal (Table 18–6) and showed that in patients with postgastric bypass hypoglycemia, the high carbohydrate meal resulted in hyperglycemia and concomitant hyperinsulinemia at about 30 minutes after a meal. Glucose levels then fell to a nadir (range 28-62 mg/dL) at 90 to 120 minutes. Eating the low carbohydrate meal, however, resulted in very little change in plasma glucose levels and only a modest increase in plasma insulin. The prolonged (5 hour) oral glucose tolerance test is not recommended for evaluation because a large number of healthy subjects will have a false-positive result. There have been case reports of insulinoma and noninsulinoma pancreatogenous hypoglycemia syndrome occurring in patients who present with hypoglycemia post-Roux-en-Y surgery. It is unclear how often this occurs. A careful history may identify those patients who have history of hypoglycemia with exercise or when meals are missed, and these individuals may require a formal 72-hour fast to rule out insulinoma.

Several different treatments can be tried for the late dumping syndrome. Dietary modification is the best option but may be difficult to sustain for some patients. More frequent meals with smaller portions of less rapidly assimilated carbohydrate and more slowly absorbed fat or protein can be tried. Alpha glucosidase inhibitors (acarbose and miglitol) can be a useful adjunct to a low carbohydrate diet in some patients. Octreotide 50 μg administered subcutaneously two or three times a day 30 minutes prior to each meal has been reported to improve symptoms in patients with severe dumping refractory to other forms of medical interventions. Information regarding long-term octreotide use is however limited. Various surgical procedures to slow down gastric emptying have been reported to improve symptoms, but long-term efficacy studies are lacking. Recently it has been reported that endoscopic gastrojejunal anastomotic reduction to induce delay in gastric pouch emptying in patients with Roux-en-Y surgery improves the dumping syndrome symptoms.

Noninsulinoma Pancreatogenous Hypoglycemia Syndrome (NIPHS)

These are patients with hyperinsulinemic hypoglycemia due to generalized islet hyperplasia. The term adult onset nesidioblastosis has also applied to this clinical entity, and it is very rare. These patients predominantly have symptoms 2 to 4 hours after meals and only rarely while fasting. As in other patients who have frequent hypoglycemia, the symptoms are predominantly those of neuroglycopenia including diplopia, dysarthria, confusion, disorientation, convulsions, and coma. At the time of hypoglycemia, these patients have elevated insulin, C-peptide and proinsulin levels and negative sulfonylurea, repaglinide, and nateglinide screens. Typically these patients have a negative 72-hour fast. Imaging studies are also negative. These patients have a positive selective arterial calcium stimulation test—usually positive for multiple arteries. NIPHS patients do not have mutations in the *KIR6.2* and *SUR1* genes, which have been abnormal in some cases of children with a syndrome of familial hyperinsulinemic hypoglycemia (discussed later) Gradient-guided partial or subtotal pancreatectomy relieves the hypoglycemic symptoms in the majority of patients.

Late Hypoglycemia of Occult Diabetes

This condition occurs in the occasional patient with impaired glucose tolerance or early type 1 or type 2 diabetes. The patient complains of hypoglycemic symptoms after consuming high carbohydrate meals. In response to an oral glucose tolerance test, they have a delayed insulin secretion pattern which produces late hypoglycemia 4 to 5 hours after ingestion of glucose. In the obese patient, treatment is directed at reduction to ideal weight. These patients often respond to reduced intake of refined sugars with multiple, spaced, small feedings that are high in dietary fiber. Those with impaired glucose tolerance should be advised to have periodic assessment for development of diabetes.

TABLE 18–6 Test meals for evaluation of postprandial hypoglycemia after Roux-en-Y gastric bypass surgery.

High Carbohydrate Meal (79% Carbohydrate, 11% Fat, and 10% Protein; 405 kcal)	Low Carbohydrate Meal (2% Carbohydrate, 74% Fat, and 24% Protein; 415 kcal)
8 oz of orange juice 1 slice of toast with 1 tsp of margarine and 2 tsp of jam	Decaffeinated black coffee or tea (without sugar) 1 egg, a 1-oz sausage patty, and a 0.5-oz slice of cheese

After overnight fast (≥8 hours) give high carbohydrate meal on day 1 and low carbohydrate meal on day 2. Measure plasma glucose and serum insulin levels before (fasting) and 30, 60, 90, 120, 150, and 180 minutes after the meal.

Data from Kellogg TA, Bantle JP, Leslie DB, et al. Postgastric bypass hyperinsulinemic hypoglycemia syndrome: characterization and response to a modified diet. *Surg Obes Relat Dis*. 2008 Jul-Aug;4(4):492-499.

Functional Alimentary Hypoglycemia

Patients present with symptoms suggestive of increased sympathetic activity—anxiety, weakness, tremor, sweating, or palpitations after meals. Physical examination and laboratory tests are normal. Previously many of these patients underwent a 5-hour oral glucose tolerance test and the detection of glucose levels in the 50s was determined to be responsible for the symptoms, and the recommendation was to modify the diet. It is now recognized that at least 10% of normal subjects who do not have any symptoms have nadir glucose levels less than 50 mg/dL during a 4- to 6-hour oral glucose tolerance test. In a study comparing responses to an oral glucose tolerance test with the response to a mixed meal test, none of the patients who had plasma glucose levels less than 50 mg/dL on oral glucose testing had low glucose values with the mixed meal. Thus, it is not recommended that these patients undergo either a prolonged oral glucose tolerance test or a mixed meal test. The patients should instead be given home blood glucose monitors (with memories) and instructed to monitor fingerstick glucose levels at the time of symptoms. Only patients who have symptoms when their fingerstick blood glucose is low (<50 mg/dL) and who have resolution of symptoms when the glucose is raised by consumption of readily absorbable carbohydrate need additional evaluation. Patients who do not have evidence for low glucose levels at the time of symptoms are generally reassured by their findings. Counseling and support should be the mainstays of therapy in this group, with dietary manipulation used only as an adjunctive form of therapy.

8. DISORDERS ASSOCIATED WITH LOW HEPATIC GLUCOSE OUTPUT

Reduced hepatic gluconeogenesis can result from a direct loss of hepatic tissue (acute yellow atrophy from fulminant viral hepatitis or toxic damage), from disorders that reduce amino acid substrate for hepatic gluconeogenesis (severe muscle wasting and inanition from anorexia nervosa, chronic starvation, uremia, and glucocorticoid deficit from adrenocortical deficiency), or from inborn errors of carbohydrate metabolism affecting glycogenolytic or gluconeogenic enzymes.

PEDIATRIC HYPOGLYCEMIA

Hypoglycemia is common in children, and prompt diagnosis and effective treatment are required to prevent seizures and brain injury in this vulnerable population. The most common cause of pediatric hypoglycemia is diabetes mellitus. Persistent hypoglycemia, however, is most often the result of a congenital defect in regulation of insulin, cortisol, or growth hormone, or an inborn error of metabolism of glucose, glycogen, or fatty acids. Pediatric hypoglycemia presents a unique challenge of distinguishing between hypoglycemia of the normal transitional glucose metabolism in the newborn and hypoglycemia that persists or presents for the first time after 3 days of life. Diagnosis of the cause of hypoglycemia relies primarily on laboratory investigations (a "critical sample") at the time of a hypoglycemic episode and Figure 18–6 shows an algorithm for elucidating a diagnosis from these results. Persistent hypoglycemia may result from: (1) a deficiency in one or more counterregulatory hormones including hypopituitarism; (2) defects in glycogenolysis, gluconeogenesis, or fatty acid oxidation; or (3) congenital hyperinsulinism.

CONGENITAL HYPERINSULINISM

As with adults, hyperinsulinism is the most frequent cause of persistent hypoglycemia in infants and children. As compared to adults, where the most common cause of hyperinsulinism is an insulin-secreting adenoma, in infants hyperinsulinism most likely stems from an underlying genetic disorder. This condition has been referenced in various ways, including persistent hyperinsulinemic hypoglycemia of infancy or islet dysregulation syndrome, but hereafter will be referred to as congenital hyperinsulinism.
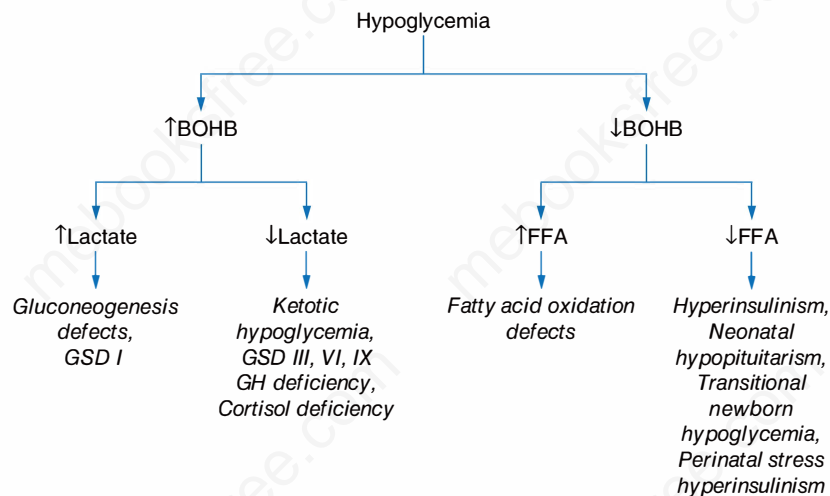
FIGURE 18–6 Algorithm to approach hypoglycemia. Shows how information from the critical sample can be used to determine major categories of hypoglycemia. BOHB, beta hydroxybutyrate; FFA, free fatty acids; GH, growth hormone.

TABLE 18–7 Types of congenital hyperinsulinism and their causes.

Transient
  Transitional neonatal hypoglycemia
  Infants of diabetic mothers
  Miscellaneous
    Small or Large for gestational age, asphyxia, and stress in infants
Persistent
  Defects in ATP-dependent potassium channel
    Sulfonylurea receptor (SUR) (ABCC8)
    Kir6.2 (KCNJ11)
  Focal vs diffuse disease
  Metabolic defects
    Glutamate dehydrogenase (GDH)
    Glucokinase (GCK)
    Hexokinase 1 (HK1)
    Pyruvate transporter (MCT1) (Exercise-induced)
    Short-chain L-3-hydroxyacyl-CoA dehydrogenase (SCHAD)
    Hepatocyte nuclear factor (HNF) 1A and HNF 4A
    Uncoupling protein 2 (UCP2)
    Phosphoglucomutase 1 (PGM1)
Syndromes
  Beckwith-Wiedemann
  Soto
  Glycosylation disorders

Transient congenital hyperinsulinism is a common disorder in the immediate neonatal period. Persistent disorders are rarer, occurring in approximately 1 in 50,000. This problem was originally attributed to an anomaly in islet development, termed **nesidioblastosis**, a reference to endocrine cell budding from pancreatic ducts. However, this budding has since been noted to represent a normal developmental process during the first year of life. Recent advances in our understanding of the regulation of insulin secretion have begun to elucidate the underlying pathophysiology of this complex disease (Table 18–7). Monogenic forms of congenital hyperinsulinism have now been attributed to mutations in eleven different genes. However, depending on the series cited, genetic mutation analysis is negative in 50% or more of cases and higher in those who are diazoxide-responsive. Timely diagnosis and aggressive treatment are essential to prevent long-term neurologic sequelae from hypoglycemia in the affected individual.

Transient Hyperinsulinism

A. Transitional neonatal hypoglycemia   It has been long recognized that plasma glucose concentrations are lower in the first few days of life in normal newborns than at older ages. Immediately after birth mean plasma glucose levels drop by 25 to 30 mg/dL and reach a nadir in the first day of life at 55 to 65 mg/dL. These glucose concentrations remain stable and are relatively unaffected by the timing of the first feeding or frequency of feedings in the first day of life. They subsequently increase over the next couple of days to reach the normal range of older children. This hypoglycemia is characterized be suppressed ketogenesis and lipolysis and a large glycemic response to intravenous glucagon injection, consistent with hyperinsulinism due to an apparent incomplete suppression of insulin secretion. This transitional neonatal hyperinsulinism may reflect a persistent reduction of the glucose threshold for insulin secretion, a normal finding in utero thought to optimize fetal growth.

B. Infants of diabetic mothers   The fetuses of mothers with poorly regulated diabetes are exposed to sustained hyperglycemia, leading to increased fetal insulin secretion, with resultant macrosomia. This increased insulin secretion persists postpartum and usually resolves after several days. As a result, infants of diabetic mothers are at high risk of developing hypoglycemia after birth. An added risk of hypoglycemia is associated with being large for gestational age at birth, mothers who required insulin during pregnancy, or mothers who were hyperglycemic during labor. They can usually be managed with early, frequent feedings or intravenous glucose until insulin secretion has normalized, often 1 to 2 days after birth.

C. Small or large for gestational age, asphyxiation, and other conditions in newborns   Various perinatal stresses are known to induce more severe and prolonged hyperinsulinism. Small-for-gestational age as well as large-for-gestational age and asphyxiated newborns, usually from toxemic mothers, frequently experience hyperinsulinism and hypoglycemia. Hyperinsulinism is also reported in erythroblastosis fetalis, sepsis, cerebral hemorrhage, prematurity, and severely stressed newborns (low Apgar scores). Hypoglycemia usually resolves within a period of several days or weeks but may persist in these settings for 6 to 12 months. Some newborns with Beckwith-Wiedemann syndrome and Soto syndrome (also known as cerebral gigantism) show β cell hyperplasia and experience transient hypoglycemia from hyperinsulinism.

Persistent Hyperinsulinism

Persistent hyperinsulinism, continuing for more than several weeks, results from a group of heterogeneous disorders, rather than a single entity, and the various subtypes are discussed below. These forms may be classified in a variety of different ways, including usual time of presentation (shortly after birth vs after months to years of life); mode of genetic transmission (autosomal recessive vs dominant); or anatomic (focal vs diffuse). In this section, the forms are grouped into defects in the pancreatic β cell K_{ATP} channel, defects affecting intracellular metabolism, and several other miscellaneous conditions or syndromes.

A. Channel defects

1. **Sulfonylurea receptor and Kir6.2**—The most common cause and the severe form of congenital hyperinsulinism appears to be related to defects in the pancreatic β cell K_{ATP} channel. This channel consists of a multimer of two proteins, the sulfonylurea receptor (SUR1), a member of the ATP-binding cassette superfamily, and Kir6.2, a member of a family of inwardly rectifying potassium channels. With increases in intracellular ATP, the channel closes, leading to membrane depolarization and insulin secretion (see Chapter 17). Inactivating mutations of either component of the channel result in a non-functional K_{ATP} channel, inappropriate β cell depolarization, and secretion of insulin, even in the face of low glucose concentrations.

The two genes, *ABCC8* and *KCNJ11*, encoding the channel components, SUR1 and Kir6.2, respectively, are located in tandem on chromosome 11p15. Numerous mutations have now been described and appear to occur more frequently in *ABCC8* than *KCNJ11*. They are usually autosomal recessive mutations, which result in complete absence of K_{ATP} channels by blocking trafficking to the plasma membrane. Dominant mutations are missense defects, which allow normal trafficking of the channel to the plasma membrane but act in a dominant negative manner to impair channel activity. The severity of the impairment manifests varying degrees of diazoxide responsiveness. Sequencing of these molecules is available commercially and may not be merely of academic interest: patients harboring some *ABCC8* or *KCNJ11* mutations do not respond well to medical therapy. Thus, knowledge of the underlying defect may influence treatment decisions for a particular patient as well as inform the family about risk to future children. Relative to other forms of congenital hyperinsulinism, patients with defects in the K_{ATP} channel often present early in life, with more marked clinical symptomatology, including macrosomia. They experience early onset and severe hypoglycemia that requires high rates of glucose infusion to normalize serum glucose concentrations and typically require pancreatectomy to restore euglycemia.

2. **Focal versus diffuse disease**—Two histologically distinct forms of congenital hyperinsulinism due to K_{ATP} channel mutations have been described, a diffuse form, mentioned above, which constitutes 35% to 70% of cases, depending on the series cited, and a focal form (focal adenomatous hyperplasia). The clinical differences between the two forms are subtle; children with the focal form are more likely to have a delay in diagnosis and hypoglycemic seizures. Diagnosis of the focal form is made by perioperative palpation and visualization by an experienced surgeon and histologic examination. [18]F DOPA positron emission tomography may also be of utility in preoperative localization of focal lesions. Histologic examination of the focal form reveals focal hyperplasia, with hypertrophied β cells harboring giant nuclei, in contrast to the diffuse form, where all the islets of Langerhans are irregular in size and contain hypertrophied β cells. The molecular explanation for the focal defect is based on a two genetic hit model, in which the child harbors a paternally inherited recessive K_{ATP} mutant allele, in either *ABCC8* or *KCNJ11*, and subsequently has an embryonic pancreas-limited chromosome recombination of 11p15.1 that results in a pancreatic lesion with paternal isodisomy (two copies of the paternally inherited recessive K_{ATP} mutant) and biallelic loss of function of K_{ATP} channel. Paternal isodisomy of the K_{ATP} mutation and the adjacent Beckwith-Weidemann syndrome (BWS) locus leads to islet overgrowth due expression and suppression of imprinted growth regulatory genes (*IGF2, H19, p57/CDKN1C*). The distinction between these two forms of hyperinsulinism has potentially important implications for therapy, because patients with focal disease are not responsive to diazoxide but may be cured with more limited partial pancreatectomy, whereas those with the diffuse form who are not responsive to diazoxide require more aggressive near-total resection, which may result in diabetes. Such patients can only be fully evaluated and treated at centers with a team of endocrinologists, interventional radiologists, pathologists, and surgeons with expertise in this disorder.

B. Metabolic defects The second most common form of congenital hyperinsulinism is the hyperinsulinism-hyperammonemia syndrome, which results from activating mutations of the glutamate dehydrogenase (*GDH*) gene. This condition is inherited in autosomal dominant fashion. This enzyme mediates oxidative deamination of glutamate to alpha ketoglutarate. Activating mutations impair GDH sensitivity to guanosine triphosphate, an allosteric inhibitor, and increase sensitivity to leucine, an allosteric activator. With increased GDH activity, increased production of alpha ketoglutarate with subsequent oxidation in the Krebs cycle generates increased ATP, which in turn activates the K_{ATP} channel and leads to membrane depolarization and insulin secretion. GDH is also expressed in the liver and kidney. Overactivity of GDH in the kidney leads to increased renal ammonia production. Thus, one hallmark of this defect is a chronic mild elevation in serum ammonia concentrations with an increase up to 3 to 5 times normal in response to a protein load. The patients with GDH mutations usually have a milder course than those with defects in the K_{ATP} channel, often presenting outside of the neonatal period. CNS manifestations include absence seizures and behavior disturbances. They often have postprandial hypoglycemia, particularly in response to higher protein loads, but may also manifest fasting hypoglycemia. They usually respond well to diazoxide, and many are able to eventually discontinue treatment.

The third most common form of congenital hyperinsulinism results from mutations in glucokinase, the first and rate-limiting step in glycolysis, and an enzyme that is considered to play an essential role in glucose sensing by the β cell. Dominant activating mutations have been described that result in increased rates of glycolysis at lower glucose concentrations. The resultant increase in the intracellular ATP/ADP ratio increases insulin secretion at any given serum glucose concentration, resulting in fasting hypoglycemia. They usually respond poorly to diazoxide. By contrast, individuals heterozygous for inactivating mutations have a form of maturity onset diabetes of the young (MODY).

Two rare forms of congenital hyperinsulinism involve the expression of genes that are not normally expressed in β cells to ensure appropriate insulin secretion. Hexokinase 1 (HK1) has a higher affinity for glucose than glucokinase, and its absence in β cells prevents insulin release at low glucose concentrations. Expression of hexokinase leads to elevations in basal insulin secretion, even at low glucose concentrations, increasing the potential for hypoglycemia. This dominant form of hyperinsulinism is responsive to diazoxide therapy. The loss of normal suppression of monocarboxylate transporter 1 (MCT1) expression in β cells leads to exercise-induced hyperinsulinism. In the setting of anaerobic exercise, the abnormal expression of MCT1 in the pancreatic beta cell allows pyruvate and lactate influx and subsequently leads to ATP generation and an inappropriate, rapid rise in insulin secretion. Therapy is largely focused on prevention with increased ingestion of carbohydrates during and following exercise.

Another autosomal recessive form of congenital hyperinsulinism results from mutations in the gene encoding the short chain L-3-hydroxyacyl-CoA dehydrogenase (SCHAD). Mutations of SCHAD cause loss of an inhibitory interaction with GDH that results in a hyperinsulinism phenotype similar to that seen with activating GDH mutations but without hyperammonemia and CNS manifestations. This recessive defect is responsive to diazoxide.

For some time, dominant inactivating mutations of the transcription factors hepatocyte nuclear factor (HNF) 1A and HNF 4A have been associated with MODY3 and MODY1, respectively. Recently, individuals bearing these mutations have been shown to manifest hyperinsulinism in early infancy, yet progress to MODY in early adulthood. The mechanism(s) underlying this transition from early hyperinsulinism and to subsequent islet failure remain unknown; animal models suggest that chronic activation of beta cells may herald later beta cell failure.

Dominant inactivating mutations of uncoupling protein 2 (UCP2) have been reported in a small number of cases of hyperinsulinism. UCP2 functions as a carrier to remove 4 carbon intermediates from the mitochondrial matrix and suppress glucose oxidation. It is assumed that reduced activity of UCP2 enhances glucose oxidation and subsequently increases insulin secretion. Phosphoglucomutase 1 (PGM1) plays an important role in regulating insulin response to glucose. Recessive inactivating mutations cause a form of glycogen storage disease that is associated with ketotic, fasting hypoglycemia, post-prandial hyperinsulinemic hypoglycemia, and abnormal protein glycosylation.

Congenital disorders in glycosylation have also been linked with congenital hyperinsulinism. The mechanism is unclear, although hypoglycosylation of the SUR with a resultant defect in trafficking to the cell membrane has been offered as a possible explanation. Nonetheless, these patients appear to be responsive to diazoxide, suggesting that at least some functional K_{ATP} channels reach the cell surface. Affected individuals have multisystem disorders, including neurologic defects.

Clinical Presentation

The cardinal symptoms of congenital hyperinsulinism are recurrent episodes of hypoglycemia and can occur from any time after birth until several years of age, or manifest even later in life, depending on the nature and severity of the defect. However, hyperinsulinism presenting in adolescence is more likely due to an insulinoma. As in adults, the symptoms of hypoglycemia are secondary to adrenergic responses or neuroglycopenia. However, in neonates and infants, symptoms are difficult to detect and may be less specific. They may include tremors, cyanosis, hypothermia, apnea or irregular breathing, lethargy, apathy, limpness, refusal to eat, high-pitched cries, and seizures of any type. Even if newborns appear asymptomatic, hypoglycemia can be severe.

Diagnosis

The differential diagnosis for pediatric hypoglycemia is quite long and includes deficiencies in counterregulatory hormones (eg, growth hormone, ACTH, glucocorticoids); defects in gluconeogenesis, glycogen synthesis, and breakdown; and disorders in fatty acid metabolism. Hyperinsulinism is suspected in hypoglycemic newborns or infants who require unusually high glucose infusion rates (12-30 mg/kg/min) to maintain blood glucose levels in the target range. Macrosomia may be another clue to hyperinsulinism, although it is not always present. The crucial diagnostic step is to obtain a *critical* blood sample for glucose, insulin, growth

hormone, cortisol, blood gases, lactate, free fatty acids, and ketones (β-hydroxybutyrate [BOHB]) during hypoglycemia (see Figure 18–6). Insulin levels must often be measured during several episodes of hypoglycemia because insulin levels at times of hypoglycemia are not always elevated to diagnostic levels. In fact, even if not elevated, insulin levels that are inappropriately measurable during episodes of hypoglycemia are still consistent with the diagnosis. As in adults, the hallmarks of hyperinsulinism include measurable plasma insulin in the face of hypoglycemia (glucose <50 mg/dL), low or unmeasurable ketones (BOHB <15 mg/dL [<1.5 mmol/L]) and free fatty acids (<28-42 mg/dL [<1.0-1.5 mmol/L]), and hyperresponsiveness to glucagon challenge, with the glycemic response to 0.5 to 1 mg (30 µg/kg) of parenteral glucagon of more than 30 mg/dL (with glucose monitored every 15 minutes for up to 45 minutes after glucagon injection).

Treatment

A. Medical therapy   Management of children with congenital hyperinsulinism remains one of the most challenging problems for pediatric endocrinologists, and affected children should be transferred to a tertiary center that has experience in managing such children. Patients may require high glucose infusion rates to maintain euglycemia, and thus a secure central line is usually necessary. They also require frequent glucose monitoring, extensive ongoing laboratory assessment to establish the underlying diagnosis, and a series of medical and/or surgical interventions to effect a cure.

Hypoglycemia in infancy has to be treated aggressively in order to prevent long-term neurologic sequelae. Relative to adults, younger children (up to age 5-6 years) appear to be especially vulnerable to such damage. Those with hyperinsulinism are particularly at risk, because ketone bodies are not present as an alternative fuel source. The therapeutic goal is to achieve glucose levels above 70 mg/dL while the infant is on an appropriate feeding schedule for age. Infants receiving appropriate therapy must be able to support a fast for at least 6 hours without hypoglycemia. In selected cases in which food refusal or intercurrent illness is a problem, hypoglycemia can be prevented by placing a gastrostomy tube to administer enteral feedings on a regular basis or with intermittent low dose intramuscular or subcutaneous glucagon injections until oral feeding resumes.

For acute management of hypoglycemia, stabilization of blood glucoses may require infusion rates of glucose up to 20 to 30 mg/kg/min, well above the 4 to 8 mg/kg/min needed to stabilize most neonates. Some centers employ aggressive enteral feeding, using frequent feeds or continuous feeding via nasogastric or gastrostomy tube, with or without cornstarch, as a means to avoid hypoglycemia. The following drugs are also used in the medical management of hyperinsulinemic hypoglycemia.

1. **Diazoxide**—Diazoxide is the drug of choice in infants who cannot be weaned from intravenous glucose. Diazoxide increases blood glucose by stabilizing the K_{ATP} channel in the open state, thereby inhibiting membrane depolarization and insulin secretion. Functionally intact SUR1 and Kir6.2 proteins are necessary for the full action of the drug, and thus, patients with channel defects often do not respond, whereas

those with other forms of hyperinsulinism do exhibit salutary responses. Diazoxide also increases catecholamine release, which suppresses insulin release and inhibits insulin's actions peripherally. The initial recommended dose is 10 to 15 mg/kg/d divided every 8 hours, up to a maximal dose of 20 mg/kg/d. Positive responses are usually seen within 48 hours if they are going to occur. Diazoxide has several important side effects that should be considered. Fluid retention can be managed by simultaneous administration of thiazide diuretics. Hypertrichosis and coarse facial changes may become quite striking and can be reduced only by decreasing the dose or discontinuing the diazoxide altogether. Hyperuricemia, leukopenia, and thrombocytopenia are rare, but routine serum studies must be monitored while therapy continues. Diazoxide is also an antihypertensive drug, but these effects are rarely encountered with oral administration.

Depending on the study cited, diazoxide appears to be efficacious in only one-fourth to one-half of the patients with hyperinsulinism and appears to be less likely to work in those patients presenting in the immediate newborn period, a population that often has more marked defects in K_{ATP} channel activity. Thus, in those patients with persistent hyperinsulinism, determining whether mutations are present in SUR1 or Kir6.2 may help anticipate the response to medical therapy and predict the need for more definitive surgical intervention.

2. **Somatostatin analogs**—These are usually a second-line approach, for those unresponsive to diazoxide. Somatostatin acts via a G protein–coupled receptor to lower intracellular calcium and to hyperpolarize the β cell membrane, thereby inhibiting insulin release. Somatostatin has a half-life of only 1 to 3 minutes, but the synthetic analog octreotide may be administered at intervals of up to 8 hours or via continuous subcutaneous infusion and is efficacious in some patients with congenital hyperinsulinism. A starting dose of 5 to 10 µg/kg/d, administered either as an intermittent bolus or via continuous subcutaneous infusion in a pump, often produces salutary initial responses, but because of tachyphylaxis due to somatostatin receptor downregulation, the dose sometimes has to be increased to as much as 40 µg/kg/d or dosed intermittently to provide a daily period of time off drug. Some physicians advocate octreotide in patients who fail to respond to diazoxide therapy alone. However, optimal control of blood glucose often cannot be achieved by adding octreotide. Nonetheless, the medication may help stabilize blood glucose concentrations in the period prior to pancreatectomy and may prove efficacious postoperatively in those patients who have persistent hyperinsulinism even with reduced β cell mass. There are some reports of long-term success (>5 years) with octreotide alone.

Short-term side effects are mostly self-limited within the first several weeks of therapy. Octreotide has nonspecific effects on the gastrointestinal tract, including decreased perfusion of the splanchnic circulation, gallbladder contractility, and bile secretion. Short-term effects may include necrotizing enterocolitis in young infants, vomiting, abdominal distension, and steatorrhea, with later risk of cholelithiasis. Possible inhibitory effects of octreotide on other hormonal axes, including effects on the pituitary somatotrope, adrenal, and thyroid, raise concerns about its long-term use, although some centers report successful and uneventful use for years without significant problems.

3. **Glucagon**—Glucagon has a place in the management of hyperinsulinism during initial stabilization of the hypoglycemic infant in the intensive care unit or prior to surgery. This agent stimulates hepatic glycogenolysis and is very effective in these patients with congenital hyperinsulinism because their glycogen stores are replete. A variety of regimens have been shown to be effective, including a bolus of 0.2 mg intravenously in cases of severe hypoglycemia followed by a continuous infusion at a dose of 2 to 10 mg/kg/h. An intramuscular glucagon injection may also be used as an emergency treatment of recurrent hypoglycemic episodes at home. Attempts have been made to administer glucagon continuously via subcutaneous infusion with limited success due to solubility issues. Newer formulations are in various development stages and show promise for sustained outpatient therapy options.

B. Surgical therapy   Pancreatectomy is undertaken when maintenance of euglycemia cannot be achieved with medical treatment alone. The extent of the pancreatectomy required depends on the form of the hyperinsulinism. Some infants may have a focal rather than diffuse form and may require only selective resection of the affected pancreatic tissue in order to effect a cure. Unfortunately, only a limited number of medical centers around the world are now equipped to conduct the pre- and perioperative evaluation to distinguish focal from diffuse disease.

The diffuse form of hyperinsulinism requires more aggressive resections. The β cell mass reduction frequently fails to achieve euglycemia, and medical therapy may have to be continued to address hyper- or hypoglycemia. If hypoglycemia persists, repeated surgical interventions may be required to remove residual and/or ectopic pancreatic tissue. Potential surgical complications include intraoperative injury to the common bile duct and adhesions with intestinal obstruction. Additional complications include exocrine pancreatic insufficiency, often requiring oral supplements at mealtimes, and diabetes mellitus.

NON-INSULIN DEPENDENT HYPOGLYCEMIA

Other remaining causes of persistent hypoglycemia in children are accounted for by deficiencies in one or more counterregulatory hormones or inborn errors of metabolism. Although several hormones are involved in the maintenance of euglycemia and produce hyperglycemic responses in excess, only deficiencies of cortisol or growth hormone have been shown to manifest hypoglycemia, which is characterized by elevated serum ketones levels. Associated physical examination findings are consistent with hypopituitarism, including microphallus, cleft lip and/or palate, neonatal jaundice, and short stature. Glycogen storage diseases (GSD) are inherited defects of glycogen synthesis or degradation. Most cases will manifest in later infancy or childhood. Ketotic hypoglycemia along with failure to thrive and hepatomegaly occur with the hepatic types of GSD (I, III, VI, IX). The most severe hepatic GSD is GSD I, which is a defect in glucose-6-phosphatase activity, the final step in glycogenolytic and gluconeogenic pathways. Lactic academia and hyperuricemia are unique to GSD I due to impaired gluconeogenesis, which is not affected in the other GSD types. Disorders of gluconeogenesis manifest hypoglycemia during episodes of significant metabolic decompensation and are associated with characteristic urine organic acid patterns

and ketosis. Fatty acid oxidation provides a significant energy source for gluconeogenesis and therefore, defects in fatty acid oxidation manifest hypoglycemia during prolonged fasting that is characterized by low ketone levels. The treatment of inborn errors of metabolism to prevent hypoglycemia is focused on decreasing the duration of fasting by frequent or continuous feeding or use of uncooked cornstarch to provide a slowly digested form of carbohydrate.

Outcome

Neurologic sequelae are the major concern with severe hypoglycemia during infancy and childhood. Multiple episodes of hypoglycemia are more often associated with sequelae than one severe hypoglycemic episode with convulsions. At least one-third of patients with congenital hyperinsulinism suffer from developmental delay based on follow-up via a telephone survey.

Patients with the familial form of congenital hyperinsulinism who harbor mutations in the SUR receptor may be at additional neurologic risk. The SUR is expressed in the brain, and defects in this molecule could potentially interfere with neural development. The role of this receptor in the brain, however, has yet to be elucidated.

Affected individuals also appear to be at higher risk for later development of diabetes. This problem may be related to reduced β cell mass following pancreatectomy or from β cell apoptosis following chronic depolarization in subjects who harbor channel defects.

REFERENCES

Amiel SA. Hypoglycemia: from the laboratory to the clinic. *Diabetes Care*. 2009;32:1364. [PMID: 19638523]

Banarer S, McGregor VP, Cryer PE. Intraislet hyperinsulinemia prevents the glucagon response to hyperglycemia despite an intact autonomic response. *Diabetes*. 2002;51:958. [PMID: 11916913]

Bolli GB, Fanelli CG. Physiology of glucose counterregulation to hypoglycemia. *Endocrinol Metab Clin North Am*. 1999;28:467. [PMID: 10500926]

Boyle PJ, Kempers SF, O'Connor AM, Nagy RJ. Brain glucose uptake and unawareness of hypoglycemia in patients with insulin-dependent diabetes mellitus. *N Engl J Med*. 1995;333:1726. [PMID: 7491135]

Cryer PE. Mechanisms of hypoglycemia-associated autonomic failure in diabetes. *N Engl J Med*. 2013;369(4):362. [PMID: 23883381]

Cryer PE, Axelrod L, Grossman A, et al. Evaluation and management of adult hypoglycemic disorders: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2009;94:709. [PMID: 19088155]

de Groot JW, Rikhof B, van Doorn J, et al. Non-islet cell tumour-induced hypoglycaemia: a review of the literature including two new cases. *Endocr Relat Cancer*. 2007;14:979. [PMID: 18045950]

de Herder WW, van Schaik E, Kwekkeboom D, et al. New therapeutic options for metastatic malignant insulinomas. *Clin Endocrinol (Oxf)*. 2011;75:277. [PMID: 21649688]

De Leon DD, Stanley CA. Mechanisms of disease: advances in diagnosis and treatment of hyperinsulinism in neonates. *Nat Clin Pract Endocrinol Metab*. 2007;3:57. [PMID: 17179930]

de Lonlay P, Fournet JC, Rahier J, et al. Somatic deletion of the imprinted 11p15 region in sporadic persistent hyperinsulinemic hypoglycemia of infancy is specific for focal adenomatous hyperplasia and endorses partial pancreatectomy. *J Clin Invest*. 1997;100:802. [PMID: 9259578]

Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study Research Group. Long-term effect of diabetes and its treatment on cognitive function. *N Engl J Med*. 2007;356:1842. [PMID: 17476010]

Doppman JL, Chang R, Fraker DL, et al. Localization of insulinomas to regions of the pancreas by intra-arterial stimulation with calcium. *Ann Intern Med*. 1995;123:269. [PMID: 7611592]

Dunne MJ, Cosgrove KE, Shepherd RM, Aynsley-Green A, Lindley KJ. Hyperinsulinism in infancy: from basic science to clinical disease. *Physiol Rev*. 2004;84:239. [PMID: 14715916]

Ekberg K, Landau BR, Wajngot A, et al. Contributions by kidney and liver to glucose production in the post-absorptive state and after 60 h of fasting. *Diabetes*. 1999;48:292. [PMID: 10334304]

Falconi M, Eriksson B, Kaltsas G, et al. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. *Neuroendocrinology*. 2016;103:153. [PMID: 26742109]

Fanelli C, Pampanelli S, Epifano L, et al. Long-term recovery from unawareness, deficient counterregulation and lack of cognitive dysfunction during hypoglycemia following institution of rational, intensive insulin therapy in IDDM. *Diabetologia*. 1994;37:1265. [PMID: 7895957]

Ferrer-García JC, Iranzo González-Cruz V, Navas-DeSolís S, et al. Management of malignant insulinoma. *Clin Transl Oncol*. 2013;15:725. [PMID: 23460559]

Glaser B, Hirsch HJ, Landau H. Persistent hyperinsulinemic hypoglycemia of infancy: long-term octreotide treatment without pancreatectomy. *J Pediatr*. 1993;123:644. [PMID: 8410523]

Grant CS. Insulinoma. *Best Pract Res Clin Gastroenterol*. 2005;19:783. [PMID: 16253900]

Hardy OT, Hernandez-Pampaloni M, Saffer JR, et al. Accuracy of [18F]fluorodopa positron emission tomography for diagnosing and localizing focal congenital hyperinsulinism. *J Clin Endocrinol Metab*. 2007;92:4706. [PMID: 17895314]

Hoe FM, Thornton PS, Wanner LA, et al. Clinical features and insulin regulation in infants with a syndrome of prolonged neonatal hyperinsulinism. *J Pediatr*. 2006;148:207. [PMID: 16492430]

Hussain K, Cosgrove KE. From congenital hyperinsulinism to diabetes mellitus: the role of pancreatic beta-cell KATP channels. *Pediatr Diabetes*. 2005;6:103. [PMID: 15963039]

Jürgensen C, Schuppan D, Neser F, et al. EUS-guided alcohol ablation of an insulinoma. *Gastrointest Endosc*. 2006;63:1059. [PMID: 16733126]

Kellogg TA, Bantle JP, Leslie DB, et al. Postgastric bypass hyperinsulinemic hypoglycemia syndrome: characterization and response to a modified diet. *Surg Obes Relat Dis*. 2008;4:492. [PMID: 18656831]

Kulke MH. Systemic therapy for advanced pancreatic neuroendocrine tumors. *Semin Oncol*. 2013;40:75. [PMID: 23391115]

Kumar U, Sasi R, Suresh S, et al. Subtype selective expression of the five somatostatin receptors (hSSTR 1–5) in human pancreatic islet cells. *Diabetes*. 1999;48:77. [PMID: 9892225]

Leibowitz G, Glaser B, Higazi AA, et al. Hyperinsulinemic hypoglycemia of infancy (nesidioblastosis) in clinical remission: high incidence of diabetes mellitus and persistent beta-cell dysfunction at long-term follow-up. *J Clin Endocrinol Metab*. 1995;80:386. [PMID: 7852494]

Levy MJ, Thompson GB, Topazian MD, et al. US-guided ethanol ablation of insulinomas: a new treatment option. *Gastrointest Endosc*. 2012;75:200. [PMID: 22078104]

Lupsa BC, Chong AY, Cochran EK, et al. Autoimmune forms of hypoglycemia. *Medicine (Baltimore)*. 2009;88:141. [PMID: 19440117]

Marks V, Teale JD. Drug-induced hypoglycemia. *Endocrinol Metab Clin North Am*. 1999;28:555. [PMID: 10500931]

Marks V, Teale JD. Hypoglycemia: factitious and felonious. *Endocrinol Metab Clin North Am*. 1999;28:579. [PMID: 10500932]

Mathur A, Gorden P, Libutti SK. Insulinoma. *Surg Clin North Am*. 2009;89:1105. [PMID: 19836487]

Meier JJ, Butler AE, Galasso R, Butler PC. Hyperinsulinemic hypoglycemia after gastric bypass surgery is not accompanied by islet hyperplasia or increased beta-cell turnover. *Diabetes Care*. 2006;29:1554. [PMID: 16801578]

Rave K, Flesch S, Kuhn-Velten WN, et al. Enhancement of blood glucose lowering effect of a sulfonylurea when coadministered with an ACE inhibitor: results of a glucose-clamp study. *Diab Metabol Res Rev*. 2005;21:459. [PMID: 15915547]

Seaquist ER, Anderson J, Childs B, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *Diabetes Care*. 2013;36:1384. [PMID: 23589542]

Seckl MJ, Mulholland PJ, Bishop AE, et al. Hypoglycemia due to an insulin-secreting small cell carcinoma of the cervix. *N Engl J Med*. 1999;341:733. [PMID: 10471459]

Service FJ. Diagnostic approach to adults with hypoglycemic disorders. *Endocrinol Metab Clin North Am*. 1999;28:519. [PMID: 10500929]

Service FJ, Natt N. The prolonged fast. *J Clin Endocrinol Metab*. 2000;85:3973. [PMID: 11095416]

Service GJ, Thompson GB, Service FJ, et al. Hyperinsulinemic hypoglycemia with nesidioblastosis after gastric-bypass surgery. *N Engl J Med*. 2005;353:249. [PMID: 16034010]

Stanley CA. Perspective on the genetics and diagnosis of congenital hyperinsulinism disorders. *J Clin Endocrinol Metab*. 2016;101:815. [PMID: 26908106]

Stumvoll M, Meyer C, Mitrakou A, Nadkarni V, Gerich JE. Renal glucose production and utilization: new aspects in humans. *Diabetologia*. 1997;40:749. [PMID: 9243094]

Tamm EP, Bhosale P, Lee JH, Rohren EM. State-of-the-art imaging of pancreatic neuroendocrine tumors. *Surg Oncol Clin N Am*. 2016;25:375. [PMID: 2701337]

Thompson GB, Service FJ, Andrews JC. Noninsulinoma pancreatogenous hypoglycemia syndrome: an update in 10 surgically treated patients. *Surgery*. 2000;128:937. [PMID: 11114627]

Thornton PS, Alter CA, Katz LE, Baker L, Stanley CA. Short- and long-term use of octreotide in the treatment of congenital hyperinsulinism. *J Pediatr*. 1993;123:637. [PMID: 8410522]

Thornton PS, Stanley CA, DeLeon DD, et al. Recommendations from the Pediatric Endocrine Society for evaluation and management of persistent hypoglycemia in neonates, infants, and children. *J Pediatr*. 2015;167:238. [PMID: 25957977]

Tucker ON, Crotty PL, Conlon KC. The management of insulinoma. *Br J Surg*. 2006;93:264. [PMID: 16498592]

Weinzimer SA, Stanley CA, Berry GT, et al. A syndrome of congenital hyperinsulinism and hyperammonemia. *J Pediatr*. 1997;130:661. [PMID: 9108870]

This page intentionally left blank

Disorders of Lipoprotein Metabolism

Mary J. Malloy, MD and John P. Kane, MD, PhD

ABCA1	ATP-binding cassette transporter A1		**IDL**	Intermediate density lipoproteins
ABCG1	ATP-binding cassette transporter G1		**Idol**	Inducible degrader of the LDL receptor
ACAT	Acyl-CoA:cholesterol acyltransferase		**LCAT**	Lecithin-cholesterol acyltransferase
ANGPTL3	Angiopoietin-like protein 3		**LDL**	Low-density lipoproteins
Apo	Apolipoprotein		**LMF1**	Lipase maturation factor 1
ARH	Autosomal recessive hypercholesterolemia		**Lp(a)**	Lipoprotein(a)
			LPL	Lipoprotein lipase
CETP	Cholesteryl ester transfer protein		**LRP-1**	LDL receptor-related protein-1
CHD	Coronary heart disease		**MCP-1**	Monocyte chemoattractant protein-1
CK	Creatine kinase		**NAFLD**	Nonalcoholic fatty liver disease
FCH	Familial combined hyperlipidemia		**NASH**	Nonalcoholic steatohepatitis
FFA	Free fatty acids		**PCPE-2**	Procollagen C-proteinase enhancer-2
FH	Familial hypercholesterolemia		**PCSK9**	Proprotein convertase subtilisin/kexin type 9
GPIHBP1	Glycosylphosphatidylinositol-anchored HDL-binding protein 1		**PDGF**	Platelet-derived growth factor
HDL	High-density lipoproteins		**SR-BI**	Scavenger receptor, class B, type I
HMG-CoA	Hydroxymethylglutaryl-CoA		**VLDL**	Very low density lipoproteins

The clinical importance of lipoprotein disorders derives chiefly from the role of lipoproteins in atherogenesis and its associated risk of coronary and peripheral vascular disease. The greatly increased risk of acute pancreatitis associated with severe hypertriglyceridemia is an additional indication for intervention. Disordered lipid metabolism is also a critical element in nonalcoholic fatty liver disease. Characterization of dyslipidemia is important for selection of appropriate treatment and may provide clues to underlying primary clinical disorders.

ATHEROSCLEROSIS

Atherosclerosis is the leading cause of death in the United States. Abundant epidemiologic evidence establishes its multifactorial character and indicates that the effects of the multiple risk factors are at least additive. Risk factors include hyperlipidemia, hypertension, smoking, diabetes, physical inactivity, decreased levels of high-density lipoproteins (HDL), hyperhomocysteinemia, inflammatory, and hypercoagulable states. Atheromas are complex lesions containing cellular elements, collagen, fibrin, and lipids. The progression of the lesion is chiefly attributable to its content of unesterified cholesterol and cholesteryl esters. Cholesterol in the atheroma originates in circulating lipoproteins. Atherogenic lipoproteins include low-density (LDL), intermediate density (IDL), very low density lipoproteins (VLDL), and Lp(a) species, all of which contain the B-100 apolipoprotein (Apo B-100). Chylomicron remnants containing apo B-48 are also atherogenic. All of these are subject to oxidation by reactive oxygen species in the tissues and also by lipoxygenases secreted by macrophages in atheromas. Oxidized lipoproteins cause impairment of endothelial cell-mediated vasodilation and stimulate endothelium to secrete

monocyte chemoattractant protein-1 (MCP-1) and adhesion molecules that recruit monocytes to the lesion. Tocopherols (vitamin E) are natural antioxidants that localize in the surface monolayers of lipoproteins, exerting resistance to oxidation. Increased oxidative stress such as that induced by smoking depletes the tocopherol content. Oxidation of lipoproteins stimulates their endocytosis via scavenger receptors on macrophages and smooth muscle cells, leading to the formation of foam cells. Recent studies strongly support a role of vitamin D in prevention of atherosclerosis, probably by influencing inflammatory activity of macrophages.

Hypertension increases access of lipoproteins to the subintima. Smoking accelerates atherogenesis by reducing HDL and increasing thrombogenesis by platelets—in addition to its pro-oxidant effect. Activated platelets release platelet-derived growth factor (PDGF), stimulating proliferation and migration of cells of smooth muscle origin into the lesion.

Activated macrophages secrete cytokines that drive an inflammatory and proliferative process. Metalloproteases secreted by macrophages weaken the atheroma so that fissuring and rupture can occur. Exposure of blood to subintimal collagen and tissue factor stimulates thrombogenesis, precipitating acute coronary events. The inverse relationship between HDL levels and atherogenesis probably reflects the role of certain species of HDL in cholesterol retrieval and in protecting lipoproteins against oxidation.

Reversal of Atherosclerosis

Angiographic intervention trials have shown that regression of atherosclerotic lesions can occur with lipid-lowering therapy. Large trials have demonstrated striking reductions in the incidence of new coronary events in individuals with hyperlipidemia who have had no prior clinical coronary disease (primary prevention) as well as in patients with antecedent disease (secondary prevention). Thus, timely hypolipidemic therapy appropriate to the lipid disorder decreases the incidence of coronary disease and reduces the need for angioplasty, stenting, and bypass surgery. Side-effects of treatment have been minimal in comparison with the magnitude of this benefit.

Average levels of LDL in the United States and northern Europe are higher than in many other nations, where the levels appear to approach the biologic norm. This probably accounts in large part for the higher incidence of coronary disease in industrialized Western nations and suggests that dietary changes that reduce levels of atherogenic lipoproteins would be beneficial.

OVERVIEW OF LIPID TRANSPORT

The Plasma Lipoproteins

Because lipids are relatively insoluble in water, they are transported in association with proteins. The simplest complexes are those formed between unesterified, or free fatty acids (FFA) and albumin, which serve to carry the FFA from peripheral adipocytes to other tissues.

Aside from some HDL species, the remainder of the lipids are transported in spherical lipoprotein complexes (Table 19–1), with core regions containing hydrophobic lipids. The principal core lipids are cholesteryl esters and triglycerides. Triglycerides predominate in the cores of chylomicrons, which transport newly absorbed lipids from the intestine, and in VLDL, which originate in liver. The relative content of cholesteryl ester is increased in the cores of remnants derived from these lipoproteins. Cholesteryl esters are the predominant core lipid in LDL and HDL. Surrounding the core in each lipoprotein is a monolayer containing amphiphilic phospholipids and unesterified (free) cholesterol. Apolipoproteins, noncovalently bound to the lipids, are located on this surface monolayer (Figure19–1).

B Apolipoproteins

Lipoproteins, excluding HDL, contain very high molecular weight B proteins that behave like intrinsic proteins of cell membranes. Unlike the smaller apolipoproteins, the B proteins do not migrate from one lipoprotein particle to another. VLDL contain the B-100 protein, which is retained in the formation of LDL from VLDL remnants by liver. Apo B-100 has a ligand domain for binding to the LDL receptor. The ligand is conformed as VLDL are transformed into LDL. Lipoprotein (a) consists of an LDL particle bound covalently to one copy of the (a) protein. The intestinal B protein, B-48, is found only in chylomicrons and their remnants.

Other Apolipoproteins

In addition to the B proteins, the following proteins are present in lipoproteins (see Table 19–1).

C apolipoproteins are lower molecular weight proteins that equilibrate rapidly among the lipoproteins. There are four distinct species: C-I, C-II, C-III, and C-IV. Apo C-II is a requisite cofactor

TABLE 19–1 Lipoproteins of human serum.

	Density Interval (g/cm^3)	Core Lipids	Diameter (nm)	Apolipoproteins in Order of Quantitative Importance
High density (HDL)	1.21-1.063	Cholesteryl ester	7.5-10.5	A-I, A-II, C, E (many others)
Low density (LDL)	1.063-1.019	Cholesteryl ester	21.5	B-100
Intermediate density (IDL)	1.019-1.006	Cholesteryl ester, triglyceride	25-30	B-100, some C and E
Very low density (VLDL)	<1.006	Triglyceride cholesteryl ester	39-100	B-100, C, E
Chylomicrons	<1.006	Triglyceride	60-500	B-48, C, E, A-I, A-II, A-IV
Lp(a)	1.04-1.08	Cholesteryl ester	21-30	B-100, (a)

FIGURE 19–1 Metabolism of chylomicrons (A-I, A-II, B-48, and C proteins, apolipoproteins; CE, cholesteryl esters; TG, triglycerides). See text for details.

for lipoprotein lipase, and Apo C-III is an inhibitor of that enzyme.

Three isoforms of Apo E—E-2, E-3, and E-4—are the products of allelic genes. Unlike Apo E-3 and Apo E-4, Apo E-2 does not contain a functional ligand for the LDL receptor. The Apo E-4 alleles are associated with early-onset Alzheimer disease and and other neurodegenerative disorders, and with increased risk of coronary disease.

Apo A-I is the major apolipoprotein of HDL. It is also present in chylomicrons and is the most abundant of the apolipoproteins of human serum (about 125 mg/dL). It is a cofactor for lecithin-cholesterol acyltransferase (LCAT). Mutations in this apolipoprotein can be amyloidogenic.

Apo A-II is an important constituent of HDL. It contains cysteine, which permits the formation of disulfide-bridged dimers with Apo E.

Apo A-IV is chiefly associated with chylomicrons.

Apo A-V plays a significant role in the removal of triglycerides from plasma.

The (a) protein is a glycoprotein that has a high degree of sequence homology with plasminogen, resulting in binding affinity for tissue plasminogen activator. It is found as a disulfide-bridged dimer with Apo B-100 in LDL-like species of lipoproteins (Lp[a] lipoproteins).

Absorption of Dietary Fat; Secretion of Chylomicrons

Dietary triglycerides are hydrolyzed in the intestine to β-monoglyceride and fatty acids by pancreatic lipase, which is activated by bile acids and a protein cofactor. The partial glycerides and fatty acids form micelles that are absorbed by intestinal epithelium. The fatty acids are reesterified with β-monoglycerides to form triglycerides, and free cholesterol is esterified with fatty

acids by acyl-CoA:cholesterol acyltransferase-2 (ACAT-2). Droplets of triglyceride with small amounts of cholesteryl esters, associated with B-48, acquire a monolayer of phospholipid and free cholesterol. Apo A-I and Apo A-II are added, and the nascent chylomicron emerges into the extracellular lymph space (see Figure 19–1). The new chylomicron begins to exchange surface components with HDL, acquiring Apo C and Apo E and losing phospholipids. This process continues as the chylomicron is carried via the intestinal lymphatics to the thoracic duct and thence into the bloodstream.

Formation of Very Low Density Lipoproteins

The liver exports triglycerides to peripheral tissues in the cores of VLDL (Figure 19–2). These triglycerides are synthesized in liver from FFA abstracted from plasma and from fatty acids synthesized *de novo.* Release of VLDL by liver is augmented by any condition that results in increased flux of FFA to liver in the absence of compensating ketogenesis. Obesity, increased caloric intake, ingestion of ethanol, and estrogens stimulate release of VLDL and are important factors in hypertriglyceridemia.

Metabolism of Triglyceride-Rich Lipoproteins in Plasma

A. Hydrolysis by lipoprotein lipase   Fatty acids derived from the triglycerides of chylomicrons and VLDL are delivered to tissues through a common pathway involving hydrolysis by the lipoprotein lipase (LPL) system. Five other gene products participate in this process. Apoproteins C-II and A-V act as cofactors for LPL, and glycosylphosphatidylinositol-anchored HDL-binding protein 1 (GPIHPB1) organizes the complex. Lipase maturation factor 1 (LMF1) is a chaperone essential for the conformation of
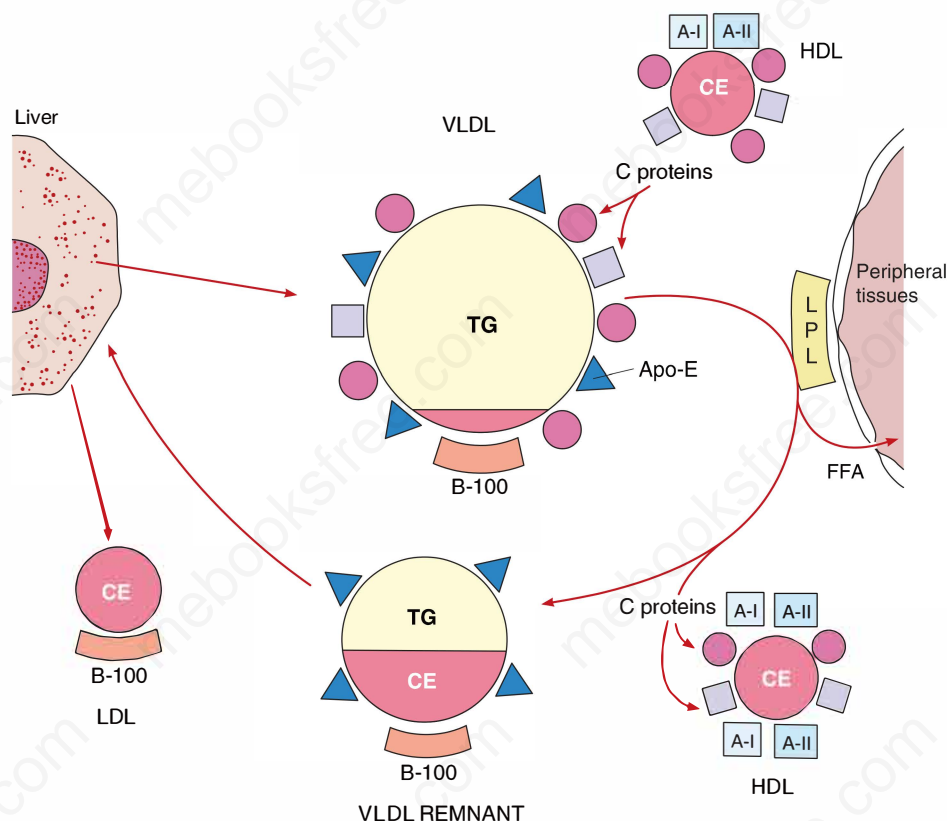
FIGURE 19–2 Metabolism of VLDL (A-I, A-II, B-48, and C proteins, apolipoproteins; CE, cholesteryl esters; TG, triglycerides). See text for details.

LPL, hepatic and endothelial lipases. Apo C-III suppresses the activity of LPL and promotes increased production of VLDL by liver. Members of a family of fatty acid transfer proteins facilitate FFA uptake in tissues. LPL is bound to capillary endothelium in heart, skeletal muscle, adipose tissue, mammary gland, and other tissues.

B. Biologic regulation of lipoprotein lipase   When glucose levels in plasma are elevated and the release of insulin is stimulated, LPL is transcriptionally upregulated in adipose tissue, and fatty acids derived from triglycerides of circulating lipoproteins are stored. During prolonged fasting, and in diabetic ketoacidosis, LPL activity of adipose tissue falls, preventing storage of fatty acids. Heparin is a cofactor for LPL. When it is given intravenously (0.1-0.2 mg/kg over 10 minutes), LPL activity is displaced into plasma, permitting its measurement.

C. Formation of lipoprotein remnants   Hydrolysis by LPL results in depletion of triglycerides in the cores of chylomicrons and VLDL, producing progressive decreases in particle diameter. Lipids from the surface and C proteins are transferred to HDL. The remnant lipoproteins thus formed contain their original complement of Apo B, significant amounts of Apo E, and little Apo C. They have lost about 70% of their triglycerides and are enriched in cholesteryl esters.

D. Fate of lipoprotein remnants   Chylomicron remnants are removed from blood quantitatively by high-affinity, receptor-mediated endocytosis in the liver. The receptors include the LDL (B-100:E) receptors and the LDL receptor-related protein-1 (LRP-1). Endocytosis of chylomicron remnants by both requires the presence of Apo E-3 or Apo E-4. The lipids enter hepatic pools, and B-48 protein is degraded. Cholesterol derived from chylomicron remnants exerts feedback control of cholesterol biosynthesis in liver.

VLDL remnants: Some are removed from blood via the B-100:E receptors (discussed later) and are degraded. Those that escape uptake are transformed into LDL. Thus, the rate of removal of VLDL remnants is a determinant of LDL production. Formation of LDL involves the removal of residual triglycerides by hepatic lipase, facilitated by Apo E. LDL contain cholesteryl esters in their cores and retain Apo B-100. In normal individuals, a major fraction of VLDL is converted to LDL, and all of the LDL Apo B comes from VLDL. In certain hypertriglyceridemic states, conversion of VLDL to LDL is decreased. In the absence of impaired conversion, increased secretion of VLDL results in increased production of LDL. This precursor-product relationship explains the clinical phenomenon referred to as the **beta shift**, an increase of LDL (beta-lipoprotein) as hypertriglyceridemia resolves. An example of this occurs temporarily following institution of insulin treatment in uncontrolled diabetes with lipemia. Insulin induces LPL activity, resulting in rapid conversion of

VLDL to LDL. Because of its longer half-life, LDL accumulates in plasma. Elevated levels of LDL may persist beyond the time when levels of triglyceride-rich lipoproteins have returned to normal. A similar phenomenon may occur when patients with familial combined hyperlipidemia (FCH) are treated with fibric acid derivatives.

E. Half-lives of lipoproteins   Normally, the half-life of chylomicrons is 5 to 20 minutes; that of VLDL is 0.5 to 1 hour; and that of LDL is about 2½ days. At triglyceride levels of 800 to 1000 mg/dL, LPL is at kinetic saturation. Increased input of triglycerides into plasma at those levels rapidly augments the hypertriglyceridemia.

F. Effect of dietary fat restriction   Individuals consuming a typical North American diet transport 75 to 100 g or more of triglyceride per day in chylomicrons, whereas the liver exports 10 to 30 g in VLDL. When LPL is saturated and triglycerides are measured in thousands of milligrams per deciliter, acute restriction of dietary fat produces a significant reduction in levels. This intervention is especially important in the lipemic patient with impending pancreatitis. If symptoms suggest that pancreatitis is imminent, oral intake should be eliminated, gastric acid should be suppressed with H_2 blockade, and the patient should not be fed by mouth until the symptoms subside and triglycerides decrease to less than 800 to 1000 mg/dL.

Catabolism of Low-Density Lipoproteins

LDL catabolism is mediated by high-affinity receptors on the cell membranes of virtually all nucleated cells but, most importantly, hepatocytes. Ligands for these LDL receptors exist in Apo B-100 and Apo E. After endocytosis, Apo B is degraded, and the receptor returns to the cell membrane. The cholesteryl esters of LDL are hydrolyzed to free cholesterol for production of cell membrane bilayers. Free cholesterol down-regulates hydroxymethylglutaryl-CoA (HMG-CoA) reductase and other rate-limiting enzymes in the biosynthetic pathway for cholesterol. Cholesterol in excess of need for membrane synthesis is esterified by ACAT-2 for storage, a process inhibited by saturated and trans fats. In addition to suppression of cholesterol biosynthesis, the entry of cholesterol via the LDL pathway leads to downregulation of LDL receptors. Further regulatory mechanisms involve shortening of the residence time of LDL receptors in cell membranes. The PCSK9 protein increases the proteolytic degradation of LDL receptors, decreasing the ability of cells to endocytose LDL. Another mechanism that leads to receptor degradation involves ubiquitination of the receptors by Idol (inducible degrader of the LDL receptor).

Metabolism of High-Density Lipoproteins

When isolated by ultracentrifugation, HDL appear to comprise two major classes: HDL_2 and HDL_3. Similar quantities of HDL_3 are isolated from serum of men and women, but about twice as much HDL_2 is found in premenopausal women. Immunochemical studies indicate that there are as many as 20 discrete species of HDL that are obscured by ultracentrifugation. One of these, the

67-kDa prebeta-1 HDL, is the primary acquisitor of cholesterol in the retrieval pathway from peripheral tissues.

A. Sources of HDL   Both liver and intestine produce HDL apolipoproteins, which organize with lipids into the native species of HDL in lymph and plasma. Excess free cholesterol and phospholipids are liberated from the surface monolayers of chylomicrons and VLDL as hydrolysis of triglycerides proceeds. Phospholipids are transferred to HDL by phospholipid transfer protein. Free cholesterol acquired by HDL is esterified by LCAT. This enzyme transfers 1 mol of fatty acid from a lecithin molecule to the hydroxyl group of unesterified cholesterol, forming cholesteryl esters. LCAT is secreted by liver. In severe hepatic parenchymal disease, levels in plasma are low and esterification of cholesterol is impeded, leading to the accumulation of free cholesterol in lipoproteins and in membranes of erythrocytes. This transforms them into the target cells classically associated with hepatic disease.

B. Metabolic roles for HDL   HDL serve as carriers for the C apolipoproteins, transferring them to nascent VLDL and chylomicrons. HDL as well as LDL deliver cholesterol to the adrenal cortex and gonads in support of steroidogenesis. HDL play a major role in the centripetal transport of cholesterol. Unesterified cholesterol, exported by the ATP-binding cassette transporter A1 (ABCA1), is acquired from the membranes of peripheral tissues by prebeta-1 HDL and is esterified by LCAT, passing through other HDL species before the cholesteryl esters are incorporated into HDL of alpha-electrophoretic mobility. Cholesterol effluxed from macrophages by the ATP-binding cassette transporter G1 (ABCG1) is acquired by larger HDL. The cholesteryl esters are then transferred to LDL and to triglyceride-rich lipoproteins mediated by cholesteryl ester transfer protein (CETP). Cholesteryl esters are also transferred from HDL to hepatocytes by scavenger receptor, class B, type I (SR-BI) receptors.

C. Catabolism of HDL   The pathways of catabolism of HDL are not yet known. Radiochemical studies indicate that Apo A-I and Apo A-II are removed from plasma synchronously and that a portion of the degradation occurs in liver and in kidney.

The Cholesterol Economy

Cholesterol is an essential constituent of the plasma membranes of cells and of myelin. It is required for adrenal and gonadal steroidogenesis and for production of bile acids by liver. Cells synthesize cholesterol, commencing with acetyl-CoA. The first committed step, mediated by HMG-CoA reductase, is the formation of mevalonic acid, which is then metabolized via a series of isoprenoid intermediaries to squalene. The latter cyclizes to form a series of sterols leading to cholesterol. A small amount of the mevalonate is converted to the isoprenoid substances ubiquinone, dolichol, and isopentenyl pyrophosphate. This pathway also yields the isoprenoid intermediaries, geranyl and farnesyl pyrophosphates, that are involved in prenylation of proteins. Prenylation provides an anchor

so that proteins such as the low molecular weight G proteins can bind to membranes. Cholesterol synthesis is tightly regulated by cholesterol or its metabolites, which down-regulate HMG-CoA reductase. Thus, cells can produce cholesterol not provided by circulating lipoproteins. Hepatocytes and intestinal epithelial cells use cholesterol for secretion of lipoproteins. In addition, cells constantly transfer cholesterol to circulating lipoproteins, chiefly HDL. Cholesterol is converted to bile acids in liver via a pathway initiated by cholesterol 7α-hydroxylase. Most of the bile acids are reabsorbed from the intestine, but the small amount that is lost in stool provides a means of elimination of cholesterol. Activity of cholesterol 7α-hydroxylase is decreased in hypothyroidism, as are the expression of LDL receptors and hepatic lipase.

At usual levels of cholesterol intake, about one-third of the amount ingested is absorbed. Most is transported to liver in chylomicron remnants, suppressing hepatic cholesterogenesis. Individuals may differ substantially in the effect of dietary cholesterol on levels of serum lipoproteins.

DIFFERENTIATION OF DISORDERS OF LIPOPROTEIN METABOLISM

Laboratory Analyses of Lipids and Lipoproteins

Chylomicrons may be secreted into plasma up to 10 hours after a meal, or longer, if alcohol is also consumed. Thus, serum lipids should be measured after a 10-hour fast. If blood glucose is not to be measured, patients may have a small non-fat meal. Because triglyceride-rich lipoproteins, their remnants, and Lipoprotein (a) are atherogenic, determination of non-HDL cholesterol by precipitation of HDL yields useful initial information.

A. Inspection   Diagnostic insights can be gained from inspection of serum or plasma before and after overnight refrigeration. Opalescence is due to light scattering by large triglyceride-rich lipoproteins. A sample begins to appear hazy when triglycerides reach 200 mg/dL (2.3 mmol/L). Chylomicrons are readily detected, because they form a white supernatant layer after refrigeration. Uncommon cases in which binding of immunoglobulins to lipoproteins takes place can be detected by the formation of a curd-like lipoprotein aggregate or a snowy precipitate as the sample cools. If one of these disorders is suspected, blood should be kept at 37°C during the formation of the clot and separation of the serum, because the critical temperatures for precipitation of the cryoglobulin complex may be higher than room temperature.

B. Laboratory techniques   Several chemical techniques provide reliable measures of cholesterol and triglycerides, an essential minimum for differentiation of disorders of lipoproteins. Unesterified and esterified cholesterol are usually measured together, so that the reported value is the total content of cholesterol in serum. A more complete characterization of lipoproteins is achieved by measurement of the cholesterol and triglyceride contents of individual lipoprotein fractions, separated by preparative ultracentrifugation. Other methods, employing magnetic resonance or electrophoresis, afford assessment of lipoprotein concentrations and particle diameters. HDL can be measured using a technique in which they are the only lipoproteins that remain in solution after treatment of the serum with heparin and manganese or other precipitants. Prognostic implications of small changes in HDL cholesterol make rigid quality controls necessary. An important determinant of the content of cholesteryl esters in HDL is the amount of triglyceride-rich lipoproteins to which the HDL are exposed in plasma. Cholesteryl esters from HDL transfer into triglyceride-rich lipoproteins, leading to an inverse dependence of HDL cholesterol on plasma triglycerides. HDL cholesterol cannot be interpreted without knowledge of the level of serum triglycerides. Elevated triglycerides can lead to erroneous diagnosis of primary HDL deficiency.

Enrichment of VLDL by cholesteryl esters is typical of familial dysbetalipoproteinemia that is due to homozygosity for Apo E-2. The Apo E genotype can be determined by PCR analysis. Clinically useful immunoassays are available for Apo B and Lp(a).

Some epidemiologic evidence suggests that LDL particles of smaller than normal diameter are associated with an increased risk of atherosclerosis. Small, dense LDL particles are a constant finding when triglyceride levels are elevated even marginally. Laboratory measurement of LDL diameters is therefore uninformative in patients with hypertriglyceridemia. LDL particle number can be assessed and elevated numbers are associated with increased risk of cardiovascular disease.

Clinical Differentiation of Abnormal Patterns of Plasma Lipoproteins

A. Preliminary screening   Epidemiologic studies in Europe and the United States have shown a progressive increase in risk of coronary disease as levels of cholesterol increase. Over the last decade, a number of objective placebo-controlled trials have demonstrated significant reduction in fatal and non-fatal outcomes with respect to coronary disease, associated with treatments that reduced the levels of atherogenic lipoproteins, chiefly LDL.

The 2013 American Heart Association/American College of Cardiology (AHA/ACC) guidelines suggest lipoprotein treatment goals based on calculated risk for individual patients. The National Lipid Association (NLA) and the International Atherosclerosis Society (IAS) have published alternate recommendations. Lipoprotein treatment goals should be individualized and based on global risk assessment that includes all the known lipoprotein and non-lipoprotein risk factors. For patients at high risk (coronary heart disease or risk equivalents, including diabetes), a suggested LDL cholesterol goal is less than 70 mg/dL (1.8 mmol/L). An LDL cholesterol goal of 100 mg/dL (2.6 mmol/dL) or below is considered appropriate for primary prevention when multiple risk factors are present. Fasting triglyceride levels above 150 mg/dL (1.7 mmol/L) merit investigation. One abnormality commonly associated with increased risk of coronary disease that is not detected if screening is limited to hyperlipidemia

is hypoalphalipoproteinemia, or deficiency of HDL. Lp(a), an important and independent risk factor, should be measured. HDL deficiency and elevated levels of Lp(a) underscore the importance of controlling other risk factors.

B. Identification of abnormal patterns   The second step in investigation of hyperlipidemia is determination of the molecular species of lipoproteins involved. In some cases, this may include multiple species; in others, qualitative properties of the lipoproteins are of diagnostic importance. Secondary hyperlipidemias of similar pattern may be the sole cause of the lipoprotein abnormality or may aggravate primary disorders of lipoprotein metabolism. The differentiation of specific primary disorders usually requires additional clinical and genetic information.

The following diagnostic protocol, based on initial measurement of cholesterol, triglycerides, and HDL in serum after a 10-hour fast, supplemented by observation of serum and by additional laboratory measurements where essential, serves as a guide to identifying abnormal patterns. The term **hyperlipidemia** denotes high levels of any class of lipoprotein; **hyperlipemia** denotes high levels of any of the triglyceride-rich lipoproteins.

Case 1: Serum Cholesterol Levels Increased; Triglycerides Normal

If the serum cholesterol level is modestly elevated (up to 260 mg/dL [6.72 mmol/L]), elevated levels of HDL may account for the observed increase. This is usually not associated with disease processes. The LDL cholesterol (in mg/dL) may be estimated by subtracting the HDL cholesterol and the estimated cholesterol contribution of VLDL from the total cholesterol level. The VLDL cholesterol is approximated as one-fifth of the serum triglyceride level (Friedewald formula).

$$\text{LDL cholesterol} = \text{total cholesterol} - \left(\frac{\text{TG}}{5} + \text{HDL cholesterol} \right)$$

Very high levels of HDL cholesterol measured by the precipitation technique can signal the presence of the abnormal lipoprotein of cholestasis (Lp-X). This disorder is characterized by elevated alkaline phosphatase activity. Rarely, deficiency of CETP or hepatic lipase can cause high levels of HDL.

Case 2: Predominant Increase of Triglycerides; Moderate Increase in Cholesterol May Be Present

Here it is apparent that the primary abnormality is an increase in triglyceride-rich VLDL or chylomicrons or both (mixed lipemia). Because both contain free cholesterol in their surface monolayers and a small amount of cholesteryl ester in their cores, the total cholesterol may be increased, although to a much smaller extent than triglycerides. Low levels of LDL often seen in hypertriglyceridemia may offset the increase in cholesterol due to the triglyceride-rich lipoproteins, especially in primary chylomicronemia.

Because VLDL and chylomicrons compete as substrates in a common removal pathway, chylomicrons are nearly always present when triglyceride levels exceed 1000 mg/dL (11.5 mmol/L).

Case 3: Cholesterol and Triglyceride Levels Both Elevated

This pattern can be the result of either of two abnormal phenotypes. One is a combined increase of VLDL and LDL. This pattern is termed combined hyperlipidemia and is one of the three phenotypic patterns encountered in kindreds with the disorder Familial Combined Hyperlipoproteinemia (FCH). The second phenotype is an increase of remnant lipoproteins derived from VLDL and chylomicrons. These particles have been partially depleted of triglyceride by LPL and enriched with cholesteryl esters by the LCAT system, such that the total content of cholesterol in serum is similar to that of triglycerides. This pattern is almost always an expression of Familial Dysbetalipoproteinemia. Diagnosis of this disorder is confirmed by a genotype demonstrating absence of the normal E-3 and E-4 alleles.

CLINICAL DESCRIPTIONS OF PRIMARY AND SECONDARY DISORDERS OF LIPOPROTEIN METABOLISM

THE HYPERTRIGLYCERIDEMIAS

Atherogenicity

Epidemiologic evidence supports the atherogenicity of VLDL, chylomicrons, and their remnants. They have been demonstrated in atherosclerotic plaques from humans. Impaired capacity of the VLDL of some individuals to accept cholesteryl esters from the LCAT reaction may also contribute to atherogenesis by impeding centripetal transport of cholesterol.

Cause of Pancreatitis

Very high levels of triglycerides in plasma are associated with a risk of acute pancreatitis, probably from the local release of FFA and lysolecithin from lipoprotein substrates in the pancreatic capillary bed. When these lipids exceed the binding capacity of albumin, they could lyse membranes of parenchymal cells, initiating a chemical pancreatitis. Many patients with lipemia have intermittent episodes of epigastric pain during which serum amylase does not reach levels commonly considered diagnostic for pancreatitis especially in patients who have had previous attacks. These episodes frequently evolve into classic pancreatitis suggesting that they represent incipient pancreatic inflammation. The progression of pancreatitis often can be prevented by rapid reduction of triglycerides, usually accomplished by restriction of all dietary fat for at least 72 hours or until symptoms disappear. In some cases, parenteral feeding, excluding fat emulsions, may be required for a few days.

Clinical Signs (Figure 19–3)

When triglyceride levels in serum exceed 3000 to 4000 mg/dL (34.5-46 mmol/L), light scattering by these particles in the blood lends a whitish cast to the venous vascular bed of the retina, a sign known as **lipemia retinalis**. Markedly elevated levels of VLDL or chylomicrons may be associated with the appearance of eruptive cutaneous xanthomas (see Figure 19–3E). These lesions, filled with foam cells, appear as yellow morbilliform eruptions 2 to 5 mm in diameter, often with erythematous areolae. They usually occur in clusters on extensor surfaces such as the elbows, knees, and buttocks. They are transient and disappear within a few weeks after triglyceride levels are reduced.

Effects of Hypertriglyceridemia on Laboratory Measurements

Light scattering from high levels of triglyceride-rich lipoproteins can cause erroneous results in most chemical determinations

FIGURE 19–3 Clinical manifestations of hyperlipidemias. **A.** Xanthelasma involving medial and lateral canthi. **B.** Severe xanthelasma and arcus corneae. **C.** Tuberous xanthomas. **D.** Large tuberous xanthoma of elbow. **E.** Eruptive xanthomas, singly and in rosettes. **F.** Xanthomas of extensor tendons of the hands. **G.** Xeroradiogram of Achilles tendon xanthoma. **H.** Xanthoma of Achilles tendon. (Normal Achilles tendons do not exceed 7 mm in diameter in the region between the calcaneus and the point at which the tendon fibers begin to radiate toward their origins.)

involving photometric measurements. Amylase activity in serum may be inhibited; therefore, lipemic specimens should be diluted for measurement of this enzyme. Because the lipoproteins are not permeable to ionic or polar molecules, their core regions constitute a second phase in plasma. When the volume of this phase becomes appreciable, electrolytes and other hydrophilic species are underestimated with respect to their true concentration in the aqueous phase. A practical rule for correcting these values is as follows: For each 1000 mg/dL (11.5 mmol/L) of triglyceride, the measured concentrations of all hydrophilic molecules and ions should be adjusted upward by 1%.

PRIMARY HYPERTRIGLYCERIDEMIA

1. DEFICIENCY OF LIPOPROTEIN LIPASE ACTIVITY

Clinical Findings

A. Symptoms and signs   The activity of LPL is dependent upon the integrity of the enzyme and its cofactor, Apo C-II. In addition, defects in Apo A-V, GPIHBP1, and LMF1 proteins can result in severe impairment of LPL activity. Because the clinical expressions of these defects are similar, they are considered together. On a typical North American diet, lipemia is usually severe (triglycerides of 2000-25,000 mg/dL) (23-287.5 mmol/L). Hepatomegaly and splenomegaly are frequently present. Foam cells laden with lipid are found in liver, spleen, and bone marrow. Splenic infarct has been described and may be a source of abdominal pain. Hypersplenism with anemia, granulocytopenia, and thrombocytopenia can occur. Recurrent epigastric pain and overt pancreatitis are frequently encountered. Eruptive xanthomas may be present. These disorders may be recognized in early infancy or may go unnoticed until an attack of acute pancreatitis occurs or lipemic serum is noted on blood sampling as late as middle age. Patients with these disorders are usually not obese and have normal carbohydrate metabolism unless pancreatitis impairs insulinogenic capacity. Estrogens intensify the lipemia by stimulating hepatic production of VLDL. Therefore, in pregnancy and lactation or during the administration of estrogenic steroids, the risk of pancreatitis increases.

B. Laboratory findings   There is a preponderance of chylomicrons in serum. Many patients have a moderate increase in VLDL, however, and in pregnant women or those receiving estrogens, a pattern of mixed lipemia is usually present. Levels of LDL are decreased, reflecting catabolism of VLDL by pathways that do not involve the production of LDL.

A presumptive diagnosis of these disorders can be made by restricting oral intake of fat to 10 to 15 g/d. Triglycerides drop precipitously, usually reaching 200 to 600 mg/dL (2.3-6.9 mmol/L) within 3 to 4 days unless the patient has significant hepatic steatosis, in which case there may be continued excretion of VLDL. Confirmation of deficiency of LPL is obtained by measurement of the lipolytic activity of plasma 10 minutes after injection of heparin (0.2 mg/kg over 10 minutes) intravenously. Analysis of lipolysis is done with or without 0.5 mol/L sodium chloride, which

inhibits LPL but does not suppress the activity of hepatic lipase. Abnormalities in LMF-1, GPIHBP, Apo A-V, and Apo C-II are identified by genotyping.

Treatment

Treatment of chylomicronemia is primarily dietary. Plasmapheresis may be helpful in the initial management of acute or incipient pancreatitis. Intake of total fat should be limited in most adult patients to 15 g/d. Because the defect involves lipolysis, both saturated and unsaturated fats must be curtailed. The diet should contain at least 5 g of polyunsaturated fat as a source of essential fatty acids, and fat-soluble vitamins must be provided. Administration of 500 mg daily of marine omega-3 fatty acids is also recommended. Adherence to this diet invariably maintains triglycerides below 1000 mg/dL (11.2 mmol/L) in the absence of pregnancy, lactation, the administration of exogenous estrogens, or alcohol. Because this is below the level at which pancreatitis usually occurs, compliant patients are at low risk. Pregnant women with these disorders require particularly close monitoring. Antisense oligonucleotide knockdown strategy for Apo C-III, currently under active investigation, may prove to be useful in management of severe hypertriglyceridemia.

2. ENDOGENOUS AND MIXED LIPEMIAS

Etiology and Pathogenesis

Endogenous lipemia (elevated VLDL) and mixed lipemia probably both result from several genetically determined disorders. Because VLDL and chylomicrons are competing substrates in the intravascular lipolytic pathway, saturating levels of VLDL impede the removal of chylomicrons. Therefore, as the severity of endogenous lipemia increases, a pattern of mixed lipemia may supervene. In other cases, the pattern of mixed lipemia appears to be present continuously. Although specific pathophysiologic mechanisms remain obscure, certain familial patterns are known. In all forms, factors that increase the rate of secretion of VLDL aggravate the hypertriglyceridemia (ie, obesity with insulin resistance, appearance of fully developed type 2 diabetes mellitus, alcohol, and exogenous estrogens). Studies of VLDL turnover indicate that either increased production or impaired removal of VLDL may be operative in different individuals. A substantial number of patients with mixed lipemia have partial defects in catabolism of triglyceride-rich lipoproteins, often due to heterozygosity for mutations in LPL. Most patients with significant endogenous or mixed lipemia have centripetal obesity.

Clinical Findings

Clinical features of these forms of hypertriglyceridemia depend on their severity and include eruptive xanthomas, lipemia retinalis, recurrent epigastric pain, and acute pancreatitis.

Endogenous lipemia is frequently an element in a constellation of metabolic abnormalities termed the **metabolic syndrome**. Insulin resistance, usually associated with central obesity, is a core feature of this syndrome. Patients may or may not have hyperglycemia. Low levels of HDL cholesterol largely reflect increased transfer of CE into circulating VLDL and, therefore, are not a

primary component of the syndrome. Hypertension and hyperuricemia are frequently present.

Treatment

The first element of treatment is severe restriction of total fat intake. The objective of long-term dietary management is reduction to ideal body weight. Because alcohol causes significant augmentation of VLDL production, abstinence is important. If weight loss is achieved, the triglycerides almost always show a marked response, often approaching normal values. When the fall in triglycerides is not satisfactory, a fibrate or nicotinic acid (in the absence of insulin resistance), singly or in combination, usually produces further reductions. When insulin resistance is present, metformin may be a useful adjunct.

3. FAMILIAL COMBINED HYPERLIPIDEMIA

Etiology

This inherited disorder, which is the most common form of hyperlipidemia, occurs in 1% to 2% of the population. The underlying process involves overproduction of VLDL. Some affected individuals have increased levels of both VLDL and LDL (combined hyperlipidemia); some have predominantly increased levels of either VLDL or LDL. The level of Apo B-100 is increased. Patterns in the serum of an individual may change with time. It is known that the offspring of an individual having any one of the three phenotypic patterns can have one of the other patterns. Affected children often have hyperlipidemia. If the child is obese, hypertriglyceridemia is likely present, whereas a child of normal weight may have only an elevated LDL.

Clinical Findings

Neither tendinous nor cutaneous xanthomas other than xanthelasma occur. This disorder appears to be inherited as a Mendelian dominant trait involving alternative loci. Modifying genes have been published, but major loci have not been identified. Factors that increase the severity of hypertriglyceridemia in other disorders aggravate the lipemia in this syndrome as well.

Treatment

The risk of coronary disease is significantly increased, and patients should be treated aggressively with diet and drugs. Because LDL levels often increase with fibrate therapy in these patients and because resins increase triglycerides, the recommended treatment is an HMG-CoA reductase inhibitor. The addition of niacin or fenofibrate may be required if triglycerides remain elevated. Fenofibrate may be used cautiously with rosuvastatin.

4. FAMILIAL DYSBETALIPOPROTEINEMIA (TYPE III HYPERLIPOPROTEINEMIA)

Etiology and Pathogenesis

A permissive genetic constitution for this disease (homozygosity for Apo E-2) occurs in about 2% of the U.S. population but the expression of hyperlipidemia apparently requires environmental and possibly additional genetic determinants. The molecular basis is the presence of isoforms of Apo E that are poor ligands for high-affinity receptors. In its fully expressed form, the lipoprotein pattern is dominated by the accumulation of remnants of VLDL and chylomicrons. Levels of LDL are decreased, reflecting interruption of the transformation of VLDL remnants to LDL. The primary defect is impaired hepatic uptake of remnants of triglyceride-rich lipoproteins. The remnant particles are enriched in cholesteryl esters such that the level of cholesterol in serum is often as high as that of triglycerides. Absence of the E-3 and E-4 alleles in genomic DNA confirms the diagnosis. Additional ligand defective mutations of Apo E are now known to result in dysbetalipoproteinemia. Some of these cause hyperlipidemia in the heterozygous state, a disorder termed **dominant dysbetalipoproteinemia**.

Clinical Findings

Hyperlipidemia and clinical signs are not usually evident before age 20. In younger patients with hyperlipidemia, hypothyroidism or obesity is likely to be present. Adults frequently have tuberous or tuberoeruptive xanthomas (see Figure 19–3C). Both tend to occur on extensor surfaces, especially elbows and knees. Tuberoeruptive xanthomas are pink or yellowish skin nodules 3 to 8 mm in diameter that often become confluent. Tuberous xanthomas— shiny reddish or orange nodules up to 3 cm or more in diameter— are usually moveable and nontender. Another type, planar xanthomas of the palmar creases, strongly suggests dysbetalipoproteinemia. The skin creases assume an orange color from deposition of carotenoids. They occasionally are raised above the level of adjacent skin and are not tender. (Planar xanthomas are also seen in cholestasis.)

Some patients have impaired glucose tolerance, which is usually associated with higher levels of blood lipids. Obesity is commonly present and tends to aggravate the lipemia. Patients with the genetic constitution for dysbetalipoproteinemia often develop severe hyperlipidemia if they are hypothyroid. Atherosclerosis of the coronary and peripheral vessels occurs with increased frequency, and the prevalence of disease of the iliac and femoral vessels is especially high.

Treatment

Management includes a weight reduction diet providing a reduced intake of cholesterol, fat, and alcohol. When the hyperlipidemia does not respond satisfactorily to diet, a fibrate or niacin in low doses (if the patient does not have insulin resistance) is usually effective. These agents can be used together in resistant cases. Some patients respond to the more potent reductase inhibitors alone, and the addition of niacin normalizes the lipid levels in most.

SECONDARY HYPERTRIGLYCERIDEMIA

1. DIABETES MELLITUS

In patients with diabetes, levels of VLDL in plasma are frequently elevated. The severe lipemia associated with absence or marked

insufficiency of insulin is attributable to decreased transcription of the *LPL* gene. The administration of insulin usually restores triglyceride levels to normal within a few days. However, if massive fatty liver is present, weeks may be required for the VLDL to return to normal while the liver secretes its triglycerides. Conversion of massive amounts of VLDL to LDL as the impedance of VLDL catabolism is relieved leads to marked accumulation of LDL that may persist for weeks, leading to a spurious diagnosis of primary hypercholesterolemia.

The moderately elevated VLDL seen in type 2 diabetes under average control probably reflects chiefly an increased flux of FFA to liver that stimulates production of triglycerides and their secretion in VLDL. In addition to VLDL, LDL levels are also somewhat increased in diabetic individuals under poor control, probably accounting in part for their increased risk of CHD. Some have much higher levels of VLDL, suggesting that an additional genetic factor predisposing to lipemia is present. Still another cause of lipemic diabetes is the compromised insulinogenic capacity that can result from acute pancreatitis. The deficiency may be severe enough to require exogenous insulin, often only in small doses. In diabetic individuals who develop nephrosis, the secondary lipemia of nephrosis compounds their hypertriglyceridemia. In the presence of hyperglycemia, lipoproteins become glycosylated, leading to their uptake by macrophages.

Lipemia may be very severe, with elevated levels of both VLDL and chylomicrons when control is poor. Lipemic patients usually have ketoacidosis when they are insulin-deficient, but lipemia can occur in its absence. Patients with type 1 diabetes who have been chronically undertreated with insulin may have mobilized most of the triglyceride from peripheral adipose tissue, so that they no longer have sufficient substrate for significant ketogenesis. These emaciated individuals may have severe lipemia and striking hepatomegaly.

In type 1 diabetes, the rigid control of blood glucose levels, which can be attained with continuous subcutaneous insulin infusion, is associated with sustained normalization of levels of both LDL and VLDL. The lipemia of type 2 diabetes usually responds well to control of the underlying disorder. In obese insulin-resistant individuals, weight loss is essential. Diets containing slowly absorbed carbohydrates are well tolerated, allowing a decrease in the burden of chylomicron triglycerides in plasma (see Chapter 17).

2. UREMIA

Uremia is associated with modest isolated increases in VLDL. The most important underlying mechanisms are probably insulin resistance and impairment of catabolism of VLDL. Many uremic patients are also nephrotic. The additional effects of nephrosis on lipoprotein metabolism may produce a combined hyperlipidemia. Patients who have had renal transplants may be receiving glucocorticoids, which induce elevation of LDL.

3. HUMAN IMMUNODEFICIENCY VIRUS INFECTION

HIV infection per se is associated with hypertriglyceridemia (see Chapter 25). A syndrome of partial lipodystrophy and insulin resistance, often with marked lipemia, occurs with multidrug treatment that includes certain inhibitors of viral proteases. Acute pancreatitis can ensue. Limited clinical experience suggests that fibric acid derivatives are of some value. Alcohol must be avoided.

4. CORTICOSTEROID EXCESS

In endogenous Cushing syndrome and with corticosteroid treatment, insulin resistance is present, and levels of LDL are increased. It appears that the hyperlipidemia is primarily due to increased secretion of VLDL, which is then catabolized to LDL. More severe lipemia ensues when steroidogenic diabetes appears, reducing catabolism of triglyceride-rich lipoproteins via the LPL pathway.

5. EXOGENOUS ESTROGENS

When estrogens are administered to normal women, triglyceride levels may increase by as much as 15%, reflecting increased production of VLDL. Paradoxically, estrogens increase the efficiency of catabolism of triglyceride-rich lipoproteins. Whereas estrogens tend to induce insulin resistance, it is not clear that this is an important mechanism, because certain nortestosterone derivatives decrease plasma triglycerides despite the induction of appreciable insulin resistance.

Certain individuals, usually those with preexisting mild lipemia, develop marked hypertriglyceridemia when receiving estrogens even in relatively small doses. Thus, triglycerides should be monitored during estrogen therapy. Contraceptive combinations with predominantly progestational effects produce less hypertriglyceridemia than purely estrogenic compounds. Transdermal delivery of estrogen probably results in lesser increases in VLDL secretion because it avoids the hepatic first-pass effect.

6. ALCOHOL INGESTION

Ingestion of appreciable amounts of alcohol does not necessarily result in significantly elevated levels of triglycerides, but many alcoholics are lipemic. Alcohol profoundly increases triglycerides in patients with primary or secondary hyperlipemias. In Zieve syndrome, the alcohol-induced lipemia is associated with hemolytic anemia and hyperbilirubinemia. Because LCAT originates in liver, severe hepatic parenchymal dysfunction may lead to deficiency in the activity of this enzyme. A resultant accumulation of unesterified cholesterol in erythrocyte membranes may account for the hemolysis seen in Zieve syndrome.

Alcohol is converted to acetate, exerting a sparing effect on the oxidation of fatty acids that are then incorporated into triglycerides. This results in hepatomegaly due to fatty infiltration and in marked enhancement of secretion of VLDL. In many individuals, there is sufficient adaptive increase in the removal capacity for triglycerides so that plasma levels are normal.

7. NONALCOHOLIC FATTY LIVER DISEASE AND NONALCOHOLIC STEATOHEPATITIS

These conditions are characterized by hepatic steatosis and abnormalities of liver enzymes in the absence of alcohol ingestion.

The cause is not yet understood, but it is likely that misdirection of fatty acids from oxidative pathways to hepatic triglyceride synthesis is involved. Five to ten percent of patients with nonalcoholic fatty liver disease (NAFLD) progress to more severe liver disease, including cirrhosis, end-stage liver disease, and hepatocellular carcinoma. In NAFLD, spontaneous improvement can occur. About 80% of patients with NAFLD or nonalcoholic steatohepatitis (NASH) have features of the metabolic syndrome, including abdominal obesity, hypertension, insulin resistance/glucose intolerance, and hypertriglyceridemia. They are at increased risk for atherosclerosis. Weight reduction with exercise and diet in overweight or obese patients is a central element in treatment. Bariatric surgery has proven effective in a large percentage of morbidly obese patients with NASH. Studies to date reveal only marginal benefit attributable to metformin. Treatment of coexisting dyslipidemia, diabetes, and hypertension is important. Statins, fenofibrate, and marine omega-3 fatty acids do not appear to increase the risk of further hepatic injury in most patients. Mixed tocopherols are beneficial.

8. NEPHROSIS

The hyperlipidemia of nephrosis is biphasic. Before serum albumin levels fall below 2 g/dL, LDL increases selectively. The synthesis and secretion of VLDL appear to be coupled to that of albumin. The increased flux of VLDL from liver increases production of LDL. As albumin falls below 1 to 2 g/dL, lipemia ensues. Impaired hydrolysis of triglycerides by LPL is due to lack of albumin as an FFA receptor. FFA bind to lipoproteins when albumin levels are low, impairing their ability to undergo lipolysis.

Because coronary disease is prevalent in patients with long-standing nephrotic syndrome, treatment of the hyperlipidemia is indicated, although few studies of the effect of treatment have been reported. The hyperlipidemia is resistant to diet. Fibrates may precipitate myopathy even in small doses. Bile acid–binding resins, niacin, and reductase inhibitors may be useful.

9. GLYCOGEN STORAGE DISEASE

In type I glycogenosis, insulin secretion is decreased, leading to increased flux of FFA to liver and increased secretion of VLDL. The low levels of insulin in plasma are the probable cause of reduced activity of LPL. The fatty liver in these patients tends to progress to cirrhosis.

Frequent small feedings, including during the night, help maintain blood glucose levels and ameliorate the lipemia. Other forms of hepatic glycogen storage disease may be associated with elevated levels of VLDL and LDL in serum.

10. HYPOPITUITARISM AND ACROMEGALY

Part of the hyperlipidemia of hypopituitarism is attributable to secondary hypothyroidism, but hypertriglyceridemia may persist after thyroxine replacement. Deficiency of growth hormone is associated with elevated levels of both LDL and VLDL. Decreased insulin levels may be the major underlying defect; however, deficiency of growth hormone may impair the disposal of FFA by oxidation and ketogenesis in the liver, favoring synthesis of triglycerides. Mild hypertriglyceridemia is often associated with acromegaly, probably resulting from insulin resistance. Although growth hormone acutely stimulates lipolysis in adipose tissue, FFA levels are normal in acromegaly.

11. HYPOTHYROIDISM

Whereas significant hypothyroidism produces elevated levels of LDL in nearly all individuals, only a few develop hypertriglyceridemia. The increase in LDL results in part from decreased conversion of cholesterol to bile acids and downregulation of LDL receptors. Lipemia, when present, is usually mild, although serum triglycerides in excess of 3000 mg/dL (34.5 mmol/L) can occur in myxedema reflecting decreased activity of hepatic lipase. Increased content of cholesteryl esters and Apo E in the triglyceride-rich lipoproteins suggests accumulation of remnant particles. Hypothyroidism often causes expression of hyperlipidemia in individuals with dysbetalipoproteinemia.

12. IMMUNOGLOBULIN-LIPOPROTEIN COMPLEX DISORDERS

Both polyclonal and monoclonal hypergammaglobulinemias may cause hypertriglyceridemia. IgG, IgM, and IgA have each been involved. Monoclonal antibodies associated with myeloma, macroglobulinemia, lymphomas, and lymphocytic leukemias have been implicated. Lupus erythematosus and other autoimmune disorders have been associated with the polyclonal type. Binding of heparin by immunoglobulin, with resulting inhibition of LPL, can cause severe mixed lipemia. More commonly, the triglyceride-rich lipoproteins have an abnormally high density, probably as a result of bound immunoglobulin, although some may be remnant-like particles. These complexes usually have gamma mobility on electrophoresis.

Xanthomatosis associated with immunoglobulin complex disease includes tuberous and eruptive xanthomas, xanthelasma, and planar xanthomas of large areas of skin. The latter are otherwise seen only in patients with cholestasis. Deposits of lipid-rich hyaline material can occur in the lamina propria of the intestine, causing malabsorption and protein-losing enteropathy. Circulating immunoglobulin-lipoprotein complexes can fix complement, leading to hypocomplementemia. In such patients, administration of whole blood or plasma can cause anaphylaxis. Hence, washed red cells or albumin are recommended when blood volume replacement is required.

Treatment is directed at the underlying disorder. The critical temperature of cryoprecipitation of some of these complexes is close to body temperature. Therefore, if plasmapheresis is indicated to remove offending immunoglobulins, it should be done at a temperature above the critical temperature measured in serum.

FAMILIAL HYPERCHOLESTEROLEMIA (FH)

LDL RECEPTOR DEFICIENCY

Etiology and Pathogenesis

This disorder occurs in approximately 1 in 500 or fewer individuals in its heterozygous form. It is a codominant trait with high penetrance. Because half of first-degree relatives may be affected, all members of a family should be screened. A selective increase in LDL exists from birth and tends to increase during childhood and adolescence such that total cholesterol in adult heterozygotes usually varies from about 250 to 400 mg/dL (6.7-10.4 mmol/L). Some individuals—especially those in kindreds in which hypertriglyceridemia is present—may have higher than normal levels of VLDL and IDL.

The underlying defect is a deficiency of normal LDL receptors on cell membranes. A number of defects in the LDL receptor gene affecting the structure, translation, modification, or transport of the receptor protein have been identified.

Some individuals have combined heterozygosity. In cases in which a kinetic mutant is combined with an ablative mutant, the hypercholesterolemia is greater than that seen in simple heterozygosity, usually in the range of 500 to 800 mg/dL (13-20.8 mmol/L). Those patients who are homozygous for null alleles have extremely severe hypercholesterolemia (approaching 1000 mg/dL [26 mmol/L] or greater) and fulminant atherosclerosis.

Production rates for LDL are moderately increased in heterozygotes and are higher in homozygotes because of increased conversion of VLDL to LDL. In the heterozygote, a greater fraction of LDL is removed by nonreceptor-dependent mechanisms than in normal subjects. In homozygotes, all removal of LDL proceeds through such pathways.

Clinical Findings

Tendon xanthomas are frequently present, causing a broadening or fusiform mass in the tendon. They can occur in almost any tendon but are most readily detected in the Achilles and patellar tendons and in the extensor tendons of the hands (see Figures 19–3F, 19–3G, and 19–3H). Patients who are physically active may complain of achillodynia. Arcus corneae (see Figure 19–3B) may occur as early as the third decade. Xanthelasma (see Figure 19–3A) may also be present. Both arcus and xanthelasma are seen in some individuals who do not have hyperlipidemia, however. Coronary atherosclerosis tends to occur prematurely in heterozygotes. It is particularly prominent in individuals who are relatively deficient in HDL or have an elevated level of Lipoprotein (a). It is probable that this represents a coincident inheritance of both traits. The homozygous form is catastrophic. Xanthomatosis progresses rapidly. Patients may have tuberous xanthomas (see Figures 19–3C and 19–3D) and elevated plaque-like xanthomas of the extremities, buttocks, interdigital webs, and aortic valves. Coronary disease may be evident in the first decade of life.

Serum cholesterol in excess of 300 mg/dL (7.8 mmol/L) in the absence of significant hypertriglyceridemia and conditions known to cause secondary hypercholesterolemia makes the diagnosis of the heterozygous disorder likely. The presence of affected first-degree relatives is supportive of this diagnosis, especially if no other phenotypes of hyperlipidemia are present in the family that would suggest FCH. The finding of tendon xanthomas is nearly pathognomonic—phytosterolemia, cerebrotendinous xanthomatosis (cholestanolosis), and autosomal recessive hypercholesterolemia (ARH) are exceptions. Although the cholesterol content of umbilical cord blood is usually elevated, the diagnosis is most easily established by measuring LDL cholesterol after the first year of life.

Treatment

Treatment of heterozygotes with HMG-CoA reductase inhibitors may normalize LDL levels. However, achieving optimal levels may require one of the binary or ternary combinations involving reductase inhibitors, niacin, bile acid sequestrants, or ezetimibe. Levels of LDL cholesterol less than 100 mg/dL (2.6 mmol/L) can be obtained with combinations of these drugs in most patients. Treatment of individuals with homozygosity or combined heterozygosity is extremely difficult. Partial control may be achieved with medications including antisense oligonucleotide directed at Apo B-100 synthesis, inhibition of microsomal triglyceride transfer protein, and ezetimibe. Statins and monoclonal antibodies directed at proprotein convertase subtilisin/kexin type 9 (PCSK9) protein are useful if some residual receptor activity is present. LDL apheresis in conjunction with medications can be very effective. Striking reduction of LDL levels is observed after liver transplantation, illustrating the important role of hepatic receptors in LDL clearance.

FAMILIAL COMBINED HYPERLIPIDEMIA (FCH)

In some individuals in kindreds with FCH (see "Primary Hypertriglyceridemia," discussed earlier), LDL are the only lipoproteins that are elevated. This pattern may vary in an individual over time. Predominant elevation of VLDL or combined elevations of LDL and VLDL may be observed in the patient's relatives. Some affected children express hyperlipidemia. In contrast to most cases of LDL receptor deficiency, the total cholesterol level may be as low as 250 mg/dL (6.5 mmol/L) and xanthomas are absent in FCH. Studies of kindreds with the FCH phenotype suggest codominant transmission. Coronary atherosclerosis is accelerated, accounting for about 15% of all coronary events in the United States. The underlying mechanism involves increased secretion of VLDL.

Treatment of the hypercholesterolemia should begin with diet and either a reductase inhibitor or niacin. It may be necessary to use a combination of these agents, or fenofibrate with a statin, to normalize levels of LDL and triglyceride.

FAMILIAL LIGAND-DEFECTIVE APO B-100

Mutations involving the ligand domain in Apo B-100 impair the ability of LDL to bind to its receptor. Two prevalent mutations at codon 3500 or 3531 occur in about one in 500 individuals and may be found in compound states with LDL receptor mutations. The hypercholesterolemia with ligand defects alone is generally less severe than in LDL receptor defects because the removal of VLDL remnants is normal, resulting in a lower production of LDL. Patients are at increased risk of coronary disease. Response to reductase inhibitors varies. Many show some resistance because upregulation of receptors cannot correct the defect completely. Receptor upregulation can decrease LDL production because IDL particles are endocytosed by liver via interaction of the LDL receptor with Apo E.

CHOLESTEROL 7α-HYDROXYLASE DEFICIENCY

Loss of function mutations in cholesterol 7α-hydroxylase result in diminished catabolism of cholesterol to bile acids and accumulation of cholesterol in hepatocytes. Downregulation of LDL receptors causes elevated LDL in plasma. VLDL may also be increased. Homozygous patients have early coronary disease, marked resistance to reductase inhibitors as single agents, and may have premature cholesterol gallstones. Heterozygous patients have moderately elevated LDL. The hyperlipidemia responds well to niacin with a reductase inhibitor.

AUTOSOMAL RECESSIVE HYPERCHOLESTEROLEMIA (ARH)

In this very rare disorder, LDL levels are markedly elevated, resulting in total cholesterol levels between 400 and 700 mg/dL (10.4-18.1 mmol/L). It is attributed to mutations in the gene for the ARH protein that is involved in endocytosis of the LDL receptor. Very large xanthomas (tuberous, tendinous, and planar) are frequently present. Arcus and xanthelasma are also common. Statins and bile acid sequestrants appear to be effective.

PROPROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 9 (PCSK9) VARIANTS

The protease PCSK9 plays a role in processing the LDL receptor. It is expressed at a high level in liver. Rare gain-of-function mutations are associated with hypercholesterolemia, due to reduction in the number of effective receptors. Hypercholesterolemia is observed in the heterozygous state.

LP(a) HYPERLIPOPROTEINEMIA

Lp(a) normally comprises a very minor fraction of circulating lipoproteins, but it may be present in high concentrations in some individuals. It contains Apo B-100 and the (a) protein, a homolog of plasminogen that can inhibit fibrinolysis. It has been demonstrated in atherosclerotic plaques, and several large studies confirm its importance as an independent risk factor for coronary disease. The Ile4399Met polymorphism is particularly associated with increased risk. Plasma levels of Lp(a) can be measured by immunoassay. Levels of Lp(a) primarily reflect genetic determinants but can also be elevated in disorders such as nephrosis because the (a) gene contains an inflammatory response element in its promoter. Niacin is effective in many but not all patients. Monoclonal antibodies directed at PCSK9 can also reduce levels of Lp(a). Decreasing plasma levels of LDL appears to reduce the risk attributed to the elevated Lp(a), at least to some extent.

SECONDARY HYPERCHOLESTEROLEMIA

HYPOTHYROIDISM

In hypothyroidism, LDL and IDL are elevated. Some patients may have lipemia, as described in the section on secondary hyperlipemia. Hyperlipidemia may occur with no overt signs or symptoms of decreased thyroid function. Biliary excretion of cholesterol and bile acids is depressed. Cholesterol stores in tissues appear to be increased, although the number of LDL receptors on cells is decreased. Activity of hepatic lipase is markedly decreased, and atherogenesis is accelerated by myxedema. The hyperlipidemia responds dramatically to treatment with thyroxine.

NEPHROSIS

As described in the section on secondary hypertriglyceridemias, nephrosis produces a biphasic hyperlipoproteinemia. The earliest alteration of lipoproteins in nephrosis is elevation of LDL. Increased secretion of VLDL by liver is probably involved. Because the lipids of the lipoprotein surfaces are altered by enrichment with sphingomyelin, lysolecithin, and FFA, the catabolism of LDL could be impaired. Perhaps the low metabolic rate in affected patients introduces metabolic changes similar to those associated with hypothyroidism. The hyperlipidemia may be an important element in the markedly increased risk of atherosclerosis in these patients. The treatment of choice is a reductase inhibitor. Ezetimibe, a bile acid–binding resin, and niacin are alternatives.

IMMUNOGLOBULIN DISORDERS

One of the lipoprotein abnormalities that can be associated with monoclonal gammopathy is elevation of LDL. A gamma lipoprotein that is a stable complex of immunoglobulin and lipoprotein may be observed on electrophoresis. Cryoprecipitation, often in the temperature range encountered in peripheral tissues when the environmental temperature is low, may occur. Patients may have symptoms from the vascular effect of complement fixation resulting from complex formation and may have hyperviscosity syndrome from the elevated immunoglobulins per se. Planar xanthomas may be present.

Treatment is directed at the underlying process. Plasmapheresis is often effective. If cryoprecipitation occurs at critical temperatures near or above room temperature, the procedure must be carried out in a special warm environment. Transfusion of whole blood or serum may be dangerous in these patients because of rapid production of anaphylatoxins from fresh complement in the serum, resulting from interaction with circulating antibody-antigen complexes. This risk can be minimized by the use of packed red blood cells and albumin in place of whole blood.

ANOREXIA NERVOSA

About 40% of patients with anorexia nervosa have elevated LDL, and levels of cholesterol may reach 400 to 600 mg/dL (10.4-15.6 mmol/L). The hyperlipidemia, which persists despite correction of hypothyroidism, is probably a result of decreased fecal excretion of bile acids and cholesterol. Serum lipoproteins return to normal when proper nutrition is restored.

CHOLESTASIS

The hyperlipidemia associated with obstruction of biliary flow is complex. Levels of cholesterol in these patients exceeding 400 mg/dL (10.4 mmol/L) usually are associated with extrahepatic obstruction or with intrahepatic tumor. Several types of abnormal lipoproteins are present. The most abundant, termed Lp-X, is a bilayer vesicle composed of unesterified cholesterol and lecithin, with associated apolipoproteins but no Apo B. Lp-X is apparent on electrophoresis as a band of zero to gamma mobility that shows metachromatic staining with Sudan black. It is these vesicular particles that cause the serum phospholipid and unesterified cholesterol content to be extremely high. Another abnormal species, called Lp-Y, contains appreciable amounts of triglycerides and Apo B. The LDL in cholestasis also contains an unusually large amount of triglycerides.

Patients may have planar xanthomas, especially at sites of minor trauma, and xanthomas of the palmar creases. Occasionally, eruptive xanthomas are present. Xanthomatous involvement of nerves may lead to symptoms of peripheral neuropathy, and the abnormal lipoproteins may be atherogenic. Whereas bilirubin levels are nearly normal in some patients with chronic cholestasis, all have elevated serum alkaline phosphatase activity.

Neuropathy or evidence of atherosclerosis is the chief indication for treatment of the hyperlipidemia. Bile acid–binding resins are of some value, whereas fibric acid derivatives may cause an increase in cholesterol. Plasmapheresis is the most effective treatment. Large doses of vitamin E are indicated to overcome severe impairment of absorption. Deficiency of other fat-soluble vitamins also occurs.

THE PRIMARY HYPOLIPIDEMIAS

Although the clinician is confronted less frequently by the problem of a striking deficiency in plasma lipids, it is important to recognize the primary and secondary hypolipidemias. A serum cholesterol less than 110 mg/dL (2.9 mmol/L) is noteworthy.

Because levels of triglycerides in normal fasting serum may be as low as 25 mg/dL (0.29 mmol/L), significance is limited to cases in which they are virtually absent.

PRIMARY HYPOLIPIDEMIA DUE TO DEFICIENCY OF HIGH-DENSITY LIPOPROTEINS

1. TANGIER DISEASE
Etiology and Pathogenesis

Severe deficiency of HDL occurs in Tangier disease. Heterozygotes lack clinical signs but have about one-half or less of the normal complement of HDL and Apo A-I in plasma. Homozygotes lack normal HDL, and Apo A-I and Apo A-II are present at extremely low levels. Serum cholesterol is usually below 120 mg/dL (3.12 mmol/L) and may be half this value. Mild hypertriglyceridemia is usually present, and LDL is greatly enriched in triglycerides. Mutations in the ATP-dependent transporter ABCA1 underlie this disorder, causing defective efflux of cholesterol from peripheral cells.

Clinical Findings

The clinical features of this rare autosomal recessive disease include large, orange-colored, lipid-filled tonsils, accumulation of cholesteryl esters in the reticuloendothelial system, and an episodic and recurrent peripheral neuropathy with predominant motor weakness in the later stages. The course of the disease is benign in early childhood, but the neuropathy may appear as early as age 8 years. Cholesteryl ester accumulates most prominently in peripheral nerve sheaths. Carotenoid pigment may be apparent in pharyngeal and rectal mucous membranes. Splenomegaly and corneal infiltration may also be present. There is some increase in risk of coronary atherosclerosis.

Treatment

Because some of the lamellar lipoprotein material in plasma is believed to originate in chylomicrons, restriction of dietary fats and cholesterol is suggested.

2. FAMILIAL HYPOALPHALIPOPROTEINEMIA
Etiology and Pathogenesis

This phenotypic pattern is a partial deficiency of HDL that may involve heterogeneous mechanisms. These presumed constitutional disorders must be differentiated from the condition in which moderately low levels of HDL are seen in individuals consuming a diet very low in fat. White and Asian men on such diets usually have HDL cholesterol levels of 38 to 42 mg/dL (1-1.1 mmol/L) by ultracentrifugal analysis, in contrast to a median value of 49 mg/dL (1.3 mmol/L) when consuming a typical North American diet. Lower levels are common in Asiatic populations and among vegetarians, where the risk of coronary disease is usually small.

HDL cholesterol must also be interpreted in the light of the amount of triglyceride-rich lipoproteins in plasma. Because cholesteryl esters are progressively transferred to the cores of triglyceride-rich lipoproteins as triglyceride levels rise, HDL cholesterol decreases as an inverse function of the triglyceride level.

Etiologic Factor in Coronary Disease

Familial hypoalphalipoproteinemia is fairly common and is an important risk factor in atherosclerosis. This abnormality may be the only apparent risk factor in many cases of premature coronary or peripheral vascular disease and accelerates the appearance of coronary disease in patients with hyperlipidemias. Hypoalphalipoproteinemia shows a strong familial incidence. Mutations in several genes have been associated, including *Apo A-I* (the principal protein of HDL), which can be amyloidogenic, *PCPE-2* (a protease involved in the maturation of Apo A-I), and the WW domain oxidoreductase (a transcriptional regulator involved in HDL metabolism). Although several mechanisms and modes of transmission may be involved, many kindreds show distributions consistent with autosomal dominance. HDL cholesterol levels are usually below 35 mg/dL (0.9 mmol/L).

Treatment

Increases in HDL cholesterol in several coronary intervention trials have been independently associated with plaque regression. Only limited means of raising HDL levels are at hand. Findings that HDL is composed of as many as 20 discrete species further complicate this problem. It is not yet known which of these species may be involved in protecting against atherosclerosis or whether their levels can be increased. Although alcohol ingestion may increase total HDL, it appears that the effect is primarily on the HDL_3 ultracentrifugal fraction, which correlates poorly with decreased risk. No recommendation for increased alcohol consumption should be made.

Heavy exercise is associated with increases in HDL in some individuals but must be approached with caution in patients who may have coronary disease. Niacin increases total HDL in many subjects, chiefly the HDL_2 ultracentrifugal fraction. Smaller increments in HDL occur with reductase inhibitors and fibric acid derivatives.

The most important reason for measuring HDL cholesterol is to identify patients who are at increased risk. Thus, just as with patients who have premature vascular disease or a family history of early arteriosclerosis, patients with low HDL should be treated more aggressively for elevated levels of the atherogenic lipoproteins. Furthermore, vigorous efforts should be directed at the control of other risk factors such as hypertension. Smoking and obesity are known to decrease HDL significantly.

3. DEFICIENCY OF LECITHIN-CHOLESTEROL ACYLTRANSFERASE

This rare autosomal recessive disorder is not expressed in clinical or biochemical form in the heterozygote. In the homozygote, clinical characteristics are variable. The primary defect is the inability to esterify free cholesterol, leading to a total deficiency of mature, spherical HDL particles. HDL species that are present include bilayer discs, small spherical particles, and prebeta-1 HDL. The diagnosis is usually made in adults, although corneal opacities may begin in childhood. Proteinuria may be an early sign. Deposits of unesterified cholesterol and phospholipid in the renal microvasculature lead to progressive loss of nephrons and ultimate renal failure. Many patients have mild to moderate normochromic anemia with target cells. Hyperbilirubinemia or peripheral neuropathy may be present. Red blood cell lipid composition is abnormal, with increased content of unesterified cholesterol and lecithin. Most patients have elevated plasma triglycerides (200-1000 mg/dL [2.3-11.2 mmol/L]), and levels of serum cholesterol vary from low normal to about 500 mg/dL (13 mmol/L), only a small fraction of which is esterified. The large triglyceride-containing lipoproteins are unusually rich in unesterified cholesterol and appear to have abnormal surface monolayers. LDL are rich in triglycerides, and abnormal vesicular lipoproteins are present in the LDL density interval. Marked restriction of dietary fat and cholesterol delays the onset of renal disease.

PRIMARY HYPOLIPIDEMIA DUE TO DEFICIENCY OF APO B–CONTAINING LIPOPROTEINS

1. RECESSIVE ABETALIPOPROTEINEMIA
Etiology and Pathogenesis

This disorder could represent a number of mutations involving the processing of Apo B or the secretion of Apo B–containing lipoproteins. The predominant cause is mutations involving the microsomal triglyceride transfer protein. Patients heterozygous for mutations have no abnormalities of lipoproteins or clinical signs. In homozygotes, no chylomicrons, VLDL, or LDL are found in plasma, leaving only HDL. Plasma triglycerides are usually less than 10 mg/dL (0.12 mmol/L) and fail to rise after a fat load. Total cholesterol is usually less than 90 mg/dL (2.3 mmol/L). There is a defect in the incorporation of newly synthesized triglycerides into chylomicrons. However, at low levels of fat intake, about 80% of the ingested triglycerides are absorbed, probably by direct absorption of fatty acids via the portal vein.

Clinical Findings

Clinical features include a paucity of adipose tissue associated with malabsorption of long-chain fatty acids due to failure of the intestine to secrete chylomicrons. Red blood cells may be acanthocytic, with a high cholesterol-phospholipid ratio. There may be progressive degeneration of the central nervous system, including cerebellar degeneration and posterior and lateral spinal tract disease. Retinal degeneration may be severe. Levels of fat-soluble vitamins in plasma may be very low. The neurologic and ophthalmologic defects are due to deficiency of vitamin E (normally transported largely in LDL). Patients are apparently normal at birth and develop steatorrhea with impaired growth in infancy.

The neuromuscular disorder often appears in late childhood with ataxia, night blindness, decreased visual acuity, and nystagmus. Cardiomyopathy with arrhythmias has been reported and may be a cause of death.

Treatment

Treatment includes administration of fat-soluble vitamins and essential fatty acids. Very large doses of mixed tocopherols (vitamin E; 100-150 IU/kg) limit the progressive central nervous system degeneration. Although vitamin A seems to correct the night blindness, it may not alter the course of retinitis pigmentosa. Vitamins D and K may also be indicated. Restriction of dietary fat minimizes steatorrhea.

2. FAMILIAL HYPOBETALIPOPROTEINEMIA

This disorder is usually attributable to defects at the Apo B locus, resulting in decreased production of protein or in the production of truncated gene products. LDL and Apo B in heterozygotes are often present at about half of normal levels. If a mutant allele resulting in the complete interdiction of Apo B synthesis is present in the homozygous state, the clinical and biochemical features may be indistinguishable from those of recessive abetalipoproteinemia, and treatment is the same as for that disorder. Very short truncations of Apo B-100 only allow the formation of abnormally dense, small LDL. Longer truncations permit the formation of larger lipoproteins, even including VLDL-like particles. Mutations in the Angiopoietin-like protein 3 (ANGPTL3) gene can cause deficiency of LDL, HDL, and triglycerides.

Clinical features may be absent in patients who produce at least low levels of LDL-like particles. However, signs and symptoms of tocopherol deficiency may be present. Treatment with mixed tocopherols (800 IU/d) is recommended for all patients.

3. CHYLOMICRON RETENTION DISEASE

This disorder presents in the neonate and appears to be based on the selective inability of intestinal epithelial cells to secrete chylomicrons due to mutations in the Secretion Associated Ras related GTPase 1B (SAR1B) gene. Affected individuals have severe malabsorption of triglycerides with steatorrhea. Levels of LDL and VLDL are about half of normal, presumably secondary to malnutrition. Tocopherol levels may be very low and may be associated with neurologic abnormalities. Clinical symptoms diminish somewhat with time if the patient is managed with a low-fat diet and tocopherol supplementation.

4. PCSK9 LOSS OF FUNCTION MUTATIONS

Loss of function mutations that result in longer residence time of the LDL receptor on the cell surface are associated with low levels of LDL cholesterol and protection from atherosclerotic vascular disease. Approximately 1 in 50 African Americans carries one of two such mutations. Other loss of function mutations are also found less commonly among Caucasians.

SECONDARY HYPOLIPIDEMIA

Hypolipidemia may be secondary to a number of diseases characterized by chronic cachexia (eg, advanced cancer). Myeloproliferative disorders can lead to extremely low levels of LDL, probably owing to increased uptake related to rapid proliferation and membrane synthesis. A wide variety of conditions leading to intestinal malabsorption produces hypolipidemia. In these situations, levels of chylomicrons, VLDL, and LDL in serum are low but never absent. Because most of the lipoprotein mass of fasting serum is of hepatic origin, massive parenchymal liver failure (eg, in Reye syndrome) can cause severe hypolipidemia. A precipitous fall in lipoprotein levels during drug treatment of hyperlipidemia can signal hepatic toxicity. Secondary hypobetalipoproteinemia occurs in oroticaciduria.

The hypolipidemias associated with immunoglobulin disorders result from diverse mechanisms. Affected patients usually have myeloma or macroglobulinemia but may have lymphomas or lymphocytic leukemia. Isolated HDL deficiency can be associated with chronic lymphocytic leukemia. Any of the major classes of immunoglobulins may be involved. In many cases, the immunoglobulins are cryoprecipitins; thus, the diagnosis may be missed if blood is not drawn and serum prepared at 37°C and observed for cryoprecipitation. Immunoglobulin-lipoprotein complexes may precipitate in various tissues. When this occurs in the lamina propria of the intestine, a syndrome of malabsorption and protein-losing enteropathy may result. Monoclonal IgA in myeloma may precipitate with lipoproteins, causing xanthomas of the gingiva and cervix. Xanthomas in the skin are usually planar and may involve intracutaneous hemorrhage, producing a classic purple xanthoma. Planar xanthomas occurring in cholestasis may be confused with this condition because the abnormal lipoprotein of cholestasis (Lp-X), like the circulating lipoprotein complex of immunoglobulin and lipoprotein, has gamma mobility on electrophoresis.

OTHER DISORDERS OF LIPOPROTEIN METABOLISM

THE LIPODYSTROPHIES

Classification

Classification of the lipodystrophies is based on their familial or acquired origin and the regional or generalized nature of the loss of adipose tissue. Mutations in a number of loci (including caveolin, peroxisome proliferator activated receptor (PPAR) gamma, lamin A/C, perilipin, seipin, and others) have been associated with various phenotypes of lipodystrophy. Insulin resistance is a common feature.

Familial generalized lipodystrophy (Seip-Berardinelli syndrome) is a rare recessive trait associated with mutations in the gene for seipin. It may be diagnosed at birth and is associated with macrosomia. Genital hypertrophy, hypertrichosis, acanthosis nigricans, hepatomegaly, insulin resistance, hypertriglyceridemia, and glucose intolerance are regularly observed.

Familial lipodystrophy of limbs and trunk (Köbberling-Dunnigan syndrome) appears to be transmitted as a dominant trait associated with mutations in the *lamin A* gene. Because sequence anomalies in other regions of the gene are associated with muscular dystrophy, cardiac conduction defects, cardiomyopathy, or axonal neuropathies, overlapping phenotypes may occur. It affects women predominantly and is not evident until puberty. The face, neck, and upper trunk are usually spared. Growth is normal, but otherwise this syndrome shares features of the generalized form noted earlier. It is frequently associated with Stein-Leventhal syndrome and often progresses to fatal cirrhosis.

Acquired forms of lipodystrophy, generalized (Lawrence syndrome) and partial (Barraquer-Simmons syndrome), usually begin in childhood, affect females predominantly, and often follow an acute febrile illness. The generalized type commonly shares the features described earlier, invariably involving the trunk and extremities but sometimes sparing the face. A sclerosing panniculitis, as seen in Weber-Christian syndrome, may appear at the outset. The partial type usually begins in the face and then involves the neck, upper limbs, and trunk. In this disorder, reduced levels of C3 complement are frequently encountered. Most patients have proteinuria, and some develop overt vascular nephritis.

Associated Disorders

Because a number of patients with disorders resembling both familial and acquired types of lipodystrophy have tumors or other lesions of the hypothalamus, appropriate neurologic evaluation should be obtained. Similarly, the physician should be alert to the association of HIV disease and collagen-vascular disorders, including scleroderma and dermatomyositis, with some cases of acquired lipodystrophy.

RARE DISORDERS

Werner Syndrome, Progeria, Infantile Hypercalcemia, Sphingolipidoses, and Niemann-Pick Disease

These disorders may be associated with hypercholesterolemia, but levels of triglycerides may be normal. Among the sphingolipidoses, HDL deficiency is typical of Gaucher disease, in which hypertriglyceridemia may also occur. Niemann-Pick disease is attributable in most cases to mutations in the *Niemann-Pick type C1* (*NPC1*) gene and may be associated with hypercholesterolemia or hypertriglyceridemia. However, a similar disorder results from mutations in the *NPC2* locus. The *NPC1* gene product is involved in the post-lysosomal transport of lipids. The neurological manifestations of Niemann-Pick disease type C are mitigated by inhibitors of glycosphingolipid synthesis.

Wolman Disease and Cholesteryl Ester Storage Disease

These recessive lipid storage disorders involve the absence and partial deficiency, respectively, of lysosomal acid lipase, resulting in abnormal cholesteryl ester and triglyceride stores in liver, spleen, adrenals, small intestine, and bone marrow. Diagnosis is established by measuring the activity of lysosomal acid lipase in plasma. Most patients have elevated levels of both LDL and VLDL. These disorders respond to the administration of recombinant lysosomal acid lipase. Statin therapy is contraindicated because it increases delivery of cholesteryl esters to hepatocytes.

Cerebrotendinous Xanthomatosis

In this recessive disorder, impaired synthesis of bile acids due to mutations in the *sterol 27-hydroxylase* gene results in increased production of cholestanol that accumulates in tissues. Plasma levels of cholestanol are elevated. Cataracts, extensive tendinous xanthomas, progressive neurologic dysfunction, and premature coronary atherosclerosis are hallmarks of this disease. Its central nervous system effects include dementia, spasticity, and ataxia. Death usually ensues before age 50 from neurologic degeneration or coronary disease. The treatment is administration of chenodeoxycholic acid.

Phytosterolemia

Mutations in the cassette half transporters ABCG5 or ABCG8 underlie this disorder, which is characterized by normal or elevated plasma cholesterol levels; high concentrations of plant sterols in serum, adipose tissue, and skin; and prominent tendinous and tuberous xanthomas. Substantially larger fractions of phytosterols and cholesterol are absorbed from the intestine than in normal individuals. Serum cholesterol levels may be as high as 700 mg/dL (18.2 mmol/L), reflecting an increase in LDL that contain phytosterol esters in addition to cholesteryl esters. Diagnosis is established by quantitation of phytosterols in plasma by gas-liquid chromatography. Premature coronary disease is common, as is polyarteritis and leukocytoclastic vasculopathy. Treatment consists of a diet restricted in plant sterols and cholesterol, and the use of ezetimibe. Bile acid–binding resins and reductase inhibitors may also be helpful.

Cholesteryl Ester Transfer Protein (CETP) Deficiency

Mutations have been identified that impair the function of CETP, resulting in the retention of cholesteryl esters in HDL. Total HDL cholesterol is increased by 30% to 50% in heterozygotes and by as much as 200 mg/dL in homozygotes. The risk of atherosclerosis may be moderately increased in homozygotes.

TREATMENT OF HYPERLIPIDEMIA

Initial therapy in all forms of hyperlipidemia is an appropriate diet. In most cases, a universal diet (discussed later) is indicated. In many subjects with lipemia or with mild hypercholesterolemia, compliance with diet is sufficient to control lipoprotein levels. Most patients with severe hypercholesterolemia or lipemia require drug therapy. Diet must be continued to achieve the full potential

of the medications. LDL cholesterol and triglycerides should be below 70 mg/dL (1.8 mmol/L) and 120 mg/dL (1.5 mmol/L), respectively, in patients with known atherosclerosis. C-reactive protein, lipoprotein-associated phospholipase A2, and other emerging biomarkers are useful in predicting risk.

Caution Regarding Drug Therapy

Women of childbearing age should be advised of the potential risk of teratogenicity of hypolipidemic agents. Treatment should be given only if pregnancy is being actively avoided. Nursing mothers should not be treated. Estrogen-containing contraceptives should be avoided or used with caution in patients with hypertriglyceridemia.

In children, hyperlipidemias other than primary hypercholesterolemias and severe cases of the metabolic syndrome may not require medication before age 16 years. Factors in the decision to start drug treatment include: the child's age and levels of lipoproteins; the severity and age at onset of symptomatic coronary disease in the child's family; and the presence in the child of other risk factors, especially diabetes or compound dyslipidemias [including hypoalphalipoproteinemia or hyper-Lp(a)lipoproteinemia in addition to a primary hypercholesterolemia]. Dietary treatment is indicated for all children with hyperlipidemia and should be started after the second year. The exception is primary chylomicronemia, in which an appropriate diet should be instituted as soon as the disease is detected.

DIETARY FACTORS IN THE MANAGEMENT OF LIPOPROTEIN DISORDERS

Restriction of Caloric Intake

The secretion of VLDL by the liver is greatly stimulated by caloric intake in excess of requirements for physical activity and basal metabolism. Therefore, the total caloric content of the diet is of greater importance than its specific composition in treating endogenous hyperlipemia. There is a positive correlation between serum levels of VLDL triglyceride and various measures of obesity, but many obese patients have normal serum lipids. On the other hand, many patients with hypertriglyceridemia—except most with chylomicronemia syndrome—are obese. As obese patients lose weight, VLDL stabilizes at lower levels. There is a modest correlation of LDL levels with body weight in the general population.

Restriction of Fat Intake

In primary chylomicronemia, all types of fats must be restricted. In the acute management of mixed lipemia with impending pancreatitis, elimination of dietary fat leads to a rapid decrease in triglycerides.

The cholesterol-lowering effect of a significant reduction in total fat is well known. It has also been shown that a 10% to 15% fall in cholesterol is achieved when individuals who have been consuming a typical North American diet restrict their intake of saturated fats to 8% of total calories. Trans-fatty acids increase

LDL cholesterol and should be avoided. Most saturated fatty acids cause increased levels of LDL cholesterol by downregulating hepatic LDL receptors. Whereas polyunsaturated fatty acids do not have this effect, they may reduce levels of HDL. Monounsaturated fatty acids increase HDL but do not increase LDL. Moderate use of monounsaturated fats such as olive oil, oleic acid–rich safflower oil, or canola oil is indicated.

Marine Omega-3 Fatty Acids

The omega-3 fatty acids are ligands for PPAR alpha. Those found in fish oils have special properties relevant to the treatment of hypertriglyceridemia and may protect against fatal arrhythmias in ischemic myocardium. Substantial decreases in triglyceride levels can be induced in some patients with severe endogenous or mixed lipemia at doses of 3 to 4 g/d of eicosapentaenoic (EPA) and docosahexaenoic (DHA) acids, combined. This should be given in divided doses with meals. EPA and DHA are potent inhibitors of platelet reactivity.

Reduction of Cholesterol Intake

The amount of cholesterol in the diet affects serum cholesterol levels, but individual responses vary. Restriction of dietary cholesterol to less than 200 mg/d (5.2 mmol/d) in normal individuals can result in a decrease of up to 10% to 15% in serum cholesterol, primarily reflecting a decrease in LDL. Dietary cholesterol and saturated or trans fat content have independent effects.

Role of Carbohydrate in Diet

When a high-carbohydrate diet is consumed, hypertriglyceridemia often develops within 48 to 72 hours, and levels of triglycerides rise to a maximum in 1 to 5 weeks. Persons with higher basal triglycerides and those consuming hypercaloric diets show the greatest effect. Because the conversion of fructose to fatty acids proceeds unregulated, it contributes disproportionally to hypertriglyceridemia compared to other saccharides. In persons with type 2 diabetes, a high-carbohydrate diet tends to increase insulin resistance. Substitution of monounsaturated fats for the carbohydrate improves insulin resistance and lipoprotein levels.

Alcohol Ingestion

Ingestion of alcohol is a common cause of secondary hypertriglyceridemia owing to overproduction of VLDL. Some individuals with familial hypertriglyceridemia are particularly sensitive to the effects of alcohol, and abstinence may normalize their triglycerides. Occasionally, chronic alcohol intake may also be associated with hypercholesterolemia. Increased cholesterol synthesis and decreased conversion to bile acids have been observed. Alcohol may account for alimentary lipemia persisting beyond 12 to 14 hours. This possibility should be excluded by the history or a repeat lipid analysis. A positive correlation has been found between alcohol intake and HDL cholesterol levels in some individuals. Because alcohol-induced changes in HDL appear primarily to involve the HDL_3 subfraction, there is no justification for the use of alcohol to increase the protective effect of HDL against atherosclerosis.

If the low HDL cholesterol is secondary to hypertriglyceridemia, alcohol must be avoided.

Antioxidants

Both alpha and gamma tocopherols (vitamin E) have recognized roles in the elimination of free radicals. Vitamin C assists this activity by restoring the tocopheroxyl radical to active tocopherol. These vitamins have been shown to restore normal vascular reactivity in hyperlipidemic patients, and some epidemiologic evidence suggests that they have an antiatherogenic effect. It is therefore reasonable to include at least 50 IU of mixed natural tocopherols and 250 mg of vitamin C in the diet each day. Larger doses may partially vitiate the effects of certain hypolipidemic drug regimens and have not been shown to have antiatherogenic potential. Selenium may also be important because it is a cofactor for one species of superoxide dismutase. Diets rich in fruits and vegetables appear to be important, providing isoflavones, quinols, and a number of carotenoid species.

B Vitamins

A significant percentage of Americans carry at least one allele for mutations in the methylene tetrahydrofolate reductase gene that diminishes its efficiency by reducing its affinity for folic acid. This, and certain other mutations, may lead to significantly elevated levels of a metabolite of methionine—homocysteine—that has toxic effects on endothelium. Supplementation with 0.4 mg of folic acid mitigates this problem. Vitamins B_6 and B_{12} participate in the metabolism of homocysteine. Thus, a B complex supplement should be used. Dietary protein should be restricted to the amount required for replacement of essential amino acids (about 0.5-1 g/kg) in patients with hyperhomocysteinemia. Oral administration of betaine, a methyl donor, can reduce homocysteine levels in severely affected patients (2 g 3 times daily). Homocysteine liberates the (a) protein from the Lp(a) complex, increasing its inhibition of fibrinolysis by plasminogen. Thus, the detection and management of hyperhomocysteinemia is important in patients with elevated Lp(a).

Other Dietary Substances

Several other nutrients have been studied in relation to atherosclerosis. Caffeine and sucrose have negligible effects on serum lipids, and their statistical relationship to coronary heart disease is generally unimpressive when data are corrected for cigarette smoking. However, when coffee is prepared by protracted boiling of the grounds, a lipid substance (cafestol) is extracted that contributes to hypercholesterolemia. Lecithin has no effect on plasma lipoproteins. A minor reduction in LDL cholesterol is associated with the addition of oat bran and certain other brans to the diet.

The Universal Diet

Dietary treatment is an important aspect of the management of all forms of lipoprotein disorders and may in some cases be all that is required. An exception is phytosterolemia in which the specific exclusion of dietary phytosterols and commercial products containing phytosterols is indicated. Knowledge of the dietary factors

mentioned earlier allows selection of appropriate modifications for an individual. However, a basic diet is useful in the treatment of most patients, the elements of which are as follows:

(1) A normal BMI should be achieved and maintained.
(2) Fat should provide less than 35% and saturated fat less than 8% of total calories. Monounsaturated oils should predominate. Olive oil contains a potent antioxidant, tyrosol that may be an important micronutrient. It is unstable when heated so should not be used for high-temperature cooking.
(3) Cholesterol should be reduced to less than 200 mg/d.
(4) Complex carbohydrates should predominate among total carbohydrates.
(5) Alcohol should be avoided in patients with hypertriglyceridemia or those requiring weight loss.
(6) Intake of trans-fatty acids should be avoided.
(7) Peroxidized fats resulting from protracted heating should be avoided.

Caloric restriction and reduction of adipose tissue mass are particularly important for patients with increased levels of VLDL and IDL. During periods of substantial weight loss, levels of VLDL and LDL tend to be lower than they would be under isocaloric conditions.

DRUGS USED IN TREATMENT OF HYPERLIPOPROTEINEMIA

BILE ACID SEQUESTRANTS

Mechanism of Action and Efficacy

Cholestyramine, colestipol, and colesevelam are cationic resins that bind bile acids in the intestinal lumen. They are not absorbed and, therefore, increase the excretion of bile acids in the stool up to tenfold. LDL levels decrease as a consequence of increased expression of high-affinity LDL receptors on hepatic cell membranes. These agents are useful only in disorders involving elevated LDL. Patients who have increased levels of VLDL may have further increases in triglycerides during treatment with resins. In combined hyperlipidemia, where the resins may be given to reduce LDL, a second agent such as fenofibrate or niacin may be required to control the hypertriglyceridemia. Levels of LDL fall 15% to 30% in compliant patients with heterozygous LDL receptor deficiency who are receiving maximal doses of the resins. These agents can contribute an incremental effect in the management of insulin resistance by an incretin-mediated mechanism. When given as a single agent, a modest reduction in new coronary events can result. When given in combined regimens with a statin and/or niacin, regression of coronary artery plaques has been demonstrated by quantitative angiography.

Drug Dosage

In disorders involving moderately high levels of LDL, 20 g of cholestyramine or colestipol daily may reduce cholesterol levels effectively. Treatment should commence at one-half this dosage to

minimize gastrointestinal side-effects. Maximum doses (30 g colestipol, 32 g cholestyramine, 3.75 g colesevelam daily) are required in more severe cases. These agents are only effective if taken with meals.

Side-Effects

Because the resins are not absorbed, systemic side-effects are absent. Patients frequently complain of a bloated sensation and constipation, both of which may be relieved by the addition of psyllium to the resin mixture. Malabsorption of fat or fat-soluble vitamins with a daily dose of resin up to 30 g occurs only in individuals with preexisting bowel disease or cholestasis. Hypoprothrombinemia has been observed in patients with malabsorption due to these causes. Cholestyramine and colestipol bind thyroxine, digitalis glycosides, and warfarin and impair the absorption of iron, thiazides, beta-blockers, ezetimibe, and other drugs. Absorption of all these (except ezetimibe) is ensured if they are administered 1 hour before or at least 2 hours after the resin. Colesevelam does not bind digoxin, warfarin, or reductase inhibitors. Because they change the composition of bile micelles, bile acid sequestrants theoretically may increase the risk of cholelithiasis, particularly in obese subjects. In practice, this risk appears to be very small. The resins should not be used as single agents in patients with hypertriglyceridemia and should be avoided in those with diverticulitis.

NIACIN (NICOTINIC ACID)

Mechanism of Action and Efficacy

Niacin (but not its amide) is able to effect major reductions in triglyceride-rich lipoproteins and, to a lesser extent, in LDL. It inhibits secretion of VLDL. It increases sterol excretion acutely, mobilizes cholesterol from tissue pools until a new steady-state is established, and decreases cholesterol biosynthesis. That it can cause a continued decrease in hepatic cholesterol production even when given with bile acid–binding resins is probably an important feature of the complementary action of these agents. Levels of HDL, particularly HDL_2, are usually significantly increased, reflecting a decrease in their fractional catabolic rate. Niacin stimulates production of tissue plasminogen activator, an effect that may be of value in preventing thrombotic events. Small, dense LDL are converted to particles of larger diameter during treatment. Niacin can reduce levels of Lp(a) significantly but variably. As a single agent, it was given in only one large intervention study which found a significant decrease in nonfatal, recurrent coronary events. In four angiographic studies in which niacin was given with a resin, a statin, or both, significant regression or decreased progression of coronary lesions was reported.

Drug Dosage

The dose of crystalline niacin required varies with the diagnosis. Optimal effect on LDL in heterozygous hypercholesterolemia is usually achieved only when 4.5 to 6 g of niacin daily is combined with a resin or reductase inhibitor. For other forms of hypercholesterolemia, dysbetalipoproteinemia, hypertriglyceridemia, and

hypoalphalipoproteinemia, 1.5 to 3.5 g/d often has a dramatic effect. Because niacin causes cutaneous flushing, it is usually started at a dosage of 100 mg 3 times daily with meals, increasing to the desired dose within 7 to 10 days. Many patients have no or only occasional flushing when stabilized on a given dose, but most must reach about 3 g/d before flushing ceases. Because the flush is prostaglandin mediated, 0.3 g of aspirin given 20 to 30 minutes before each dose when treatment is initiated or the dose increased (or equivalent doses of other cyclooxygenase inhibitors) may mitigate this symptom. It is important to counsel the patient that the flushing is a harmless cutaneous vasodilation and that the drug should be taken with meals 2 or 3 times daily. A daily dose of 6.5 g is the maximum under any circumstances. Niaspan, a prescription extended-release niacin, appears to be safe in doses up to 2 g daily.

Side-Effects

Some patients have reversible elevations of serum glutamic aminotransferase or alkaline phosphatase activities up to 3 times the upper limit of normal that do not appear to be clinically significant. In a group of patients treated continuously for up to 15 years, no significant liver disease developed despite such enzyme abnormalities. Rarely, patients develop a chemical hepatitis signaled by malaise, anorexia, and nausea. Aminotransferase levels are significantly elevated, and levels of lipoproteins may fall precipitously. Treatment should be stopped immediately. About one-fifth of patients have mild hyperuricemia that tends to be asymptomatic unless the patient has had gout. In such cases, allopurinol can be added to the regimen. Some patients have moderate elevations of blood glucose during treatment. Again, this is reversible except in those who have latent type 2 diabetes. The use of niacin should be considered carefully in patients with insulin resistance. Diabetic patients receiving insulin can often be treated successfully. A more common side-effect is gastric irritation, which responds well to H_2 blockers and antacids. Antacids that contain aluminum should be avoided. Rarely, patients develop acanthosis nigricans, which usually clears if the drug is discontinued. Some patients can have cardiac arrhythmias, mostly atrial, while taking niacin. Reversible macular edema has been described rarely. A principal metabolite of niacin is N-methylnicotinamide, the formation of which consumes methyl groups resulting in the production of macrocytic erythrocytes that are without apparent inimical effect.

Niacin should be avoided in patients with peptic ulcer or hepatic parenchymal disease. Liver function, uric acid, and blood glucose should be evaluated before commencing treatment and periodically thereafter.

Most over-the-counter sustained-release preparations should be avoided because of the risk of fulminant hepatic failure. However, if the daily dose is limited to 2 g or less, this rare consequence is probably unlikely.

FIBRIC ACID DERIVATIVES

Mechanism of Action and Efficacy

Gemfibrozil and fenofibrate, which are ligands for PPAR alpha, decrease lipolysis in adipose tissue, reduce levels of circulating

triglycerides, and cause modest reductions in LDL. However, in some patients, reductions in VLDL levels are attended by increased LDL levels. They cause moderate increases in levels of HDL, including the protein moiety. One primary prevention, and one secondary prevention trial using gemfibrozil as monotherapy showed reduction in new coronary events.

Drug Dosage

The fibrates may be useful in the treatment of patients with severe endogenous lipemia, familial dysbetalipoproteinemia, and some patients with combined hyperlipidemia. The usual dose of gemfibrozil is 600 mg twice daily. That of fenofibrate varies with the preparation.

Side-Effects

Skin eruptions, gastrointestinal symptoms, and muscle symptoms have been described as well as blood dyscrasias and elevated levels of aminotransferases and alkaline phosphatase. These drugs enhance the effects of the coumarin and indanedione anticoagulants and increase lithogenicity of bile. Concomitant use of fibrates with certain reductase inhibitors increases the risk of myopathy. Fenofibrate is most compatible with pravastatin or rosuvastatin with respect to cytochrome metabolism. Fibrates should be avoided during pregnancy and lactation and are contraindicated if renal or severe hepatic disease is present.

HMG-CoA REDUCTASE INHIBITORS (TABLE 19–2)

Mechanism of Action and Efficacy

Several closely related structural analogs of HMG-CoA act as competitive inhibitors of HMG-CoA reductase, a key enzyme in the cholesterol biosynthetic pathway. Of these, lovastatin, pravastatin, simvastatin, fluvastatin, atorvastatin, pitavastatin, and rosuvastatin are approved for use in the United States. Inhibition of cholesterol biosynthesis induces an increase in high-affinity LDL receptors in the liver, increasing removal of LDL from plasma and decreasing production of LDL. Modest increases in HDL cholesterol and limited decreases in VLDL levels can be achieved. These drugs have no appreciable effect in patients with severe

TABLE 19–2   Reductase inhibitors.

Drugs (in Increasing Order of Potency)	Dose Range (mg)
Fluvastatin	20-80
Pravastatin	10-80
Lovastatin	10-80
Simvastatin	5-80
Atorvastatin	5-80
Resuvastatin	5-40
Pitavastatin	1-4

hypertriglyceridemia. Some of the cholesterol-independent effects of reductase inhibitors appear to involve enhanced stability of atherosclerotic lesions and decreased oxidative stress and vascular inflammation, with improved endothelial function. Institution of treatment with a reductase inhibitor should begin immediately in all patients with acute coronary syndromes regardless of cholesterol level. A number of primary and secondary trials with statins as monotherapy have demonstrated significant reduction in endpoints including primary or recurrent myocardial infarction, coronary revascularization, atherosclerotic stroke, or cardiac death. Reduction of LDL cholesterol levels to or below currently accepted goals produced an approximate 35% reduction in new coronary events over the 3 to 5 year duration of the trials. Longer term follow-up suggests that the impact on new coronary events will deepen over time, consistent with observations on individuals with familial hypobetalipoproteinemia or loss of function variants in the *PCSK9* gene.

Drug Dosage

These drugs are the most effective individual oral agents for treatment of hypercholesterolemia. Their effects are amplified significantly when combined with niacin, ezetimibe, or resin. Daily dosage ranges are presented in Table 19–2. Patients with LDL receptor defects usually require higher doses. Because information on long-term safety is lacking, use of these agents in children should be restricted to those with homozygous hypercholesterolemia and selected heterozygotes who are at particularly high risk. Women who are lactating, pregnant, or likely to become pregnant should not be given these drugs.

Side-Effects

These agents are generally well tolerated. Side-effects, often transient, include rashes and changes in bowel function. Myopathy with markedly elevated creatine kinase (CK) levels occurs infrequently and is often associated with mutations in the Solute Carrier Organic anion transporter family, member 1B1 (SLCO1B1) gene. Rarely, myopathy can progress to rhabdomyolysis with myoglobinuria and renal shutdown. Among the rare hypersensitivity syndromes (Lupus-like syndromes and dermatomysitis) associated with statin treatment, is a more recently described autoimmune myopathy. This presents as severe proximal muscle weakness and pain that does not diminish after the statin is discontinued, is HMGCoA reductase antibody positive, and requires immunosuppressive treatment, usually with good results. There is an increased incidence of myopathy in patients receiving certain reductase inhibitors with cyclosporine, fibric acid derivatives, macrolides, HIV protease inhibitors, nefazodone, verapamil, and ketoconazole. Other drugs that compete for metabolism by cytochrome P450 3A4 can be expected to have the same effect. Because pravastatin and rosuvastatin do not compete with these agents for metabolism by cytochrome P450 enzymes, they appear to be compatible with them. Fluvastatin is chiefly metabolized by cytochrome P450 2C9. Thus, competitors for that pathway may cause accumulation of this reductase inhibitor. Moderate myopathy is usually rapidly reversible on cessation of statin therapy.

Minor elevations of CK activity in plasma are noted more frequently, especially with unusual physical activity. CK levels should be measured before starting therapy and monitored at regular intervals in selected patients. Older patients, those taking higher doses and multiple other drugs, those who have diabetes or renal insufficiency, and those with baseline elevated CK levels should be observed more frequently. A form of statin-associated myopathy can occur in the absence of CK elevations.

Moderate, often intermittent elevations of serum aminotransferases (up to 3 times normal) occur in some patients. If the patient is asymptomatic, therapy may be continued if activity is measured frequently (at 1- to 2-month intervals) and the levels are stable. In about 2% of patients, some of whom have underlying liver disease or a history of alcohol use, aminotransferase activity may exceed 3 times the normal limit. This can occur after 3 or more months of continuous therapy and may portend more severe hepatic toxicity such as that described in the section on niacin. The reductase inhibitor should be discontinued promptly in these patients. These agents are contraindicated in the presence of active liver disease and should be used with caution in patients with a history of liver disease. They should be discontinued temporarily during hospitalization for major surgery.

Large scale intervention trials have revealed an increased incidence of type 2 diabetes in individuals treated with statins compared with those receiving placebo. A meta-analysis revealed the increment of risk to be 9%. Patients who developed diabetes derived the same benefit in cardiovascular outcome as those receiving statins who did not develop diabetes. This effect was greater among older subjects. Periodic assessment of HgbA1c is indicated in patients receiving statins.

CHOLESTEROL ABSORPTION INHIBITORS

Mechanism of Action and Efficacy

Ezetimibe, the first of this class, inhibits the absorption of cholesterol and phytosterols by enterocytes. By interrupting the enterohepatic circulation of sterols secreted in bile, it increases sterol elimination. Recent trials indicate that ezetimibe contributes to reduction of new coronary events.

Drug Dosage

Ezetimibe is useful in treating primary hypercholesterolemias and phytosterolemia. Concomitant use of fibric acid derivatives can increase the blood concentration of this drug. Resins can decrease its absorption. It should be avoided in pregnant or lactating women and in patients with liver disease, and used with caution in patients receiving cyclosporine. A dose of 10 mg daily reduces cholesterol by 15% to 20%.

Side-Effects

Very few side-effects have been reported. The prevalence of elevated liver enzymes may be modestly increased when ezetimibe is given with a reductase inhibitor. Biopsy-proven myopathy has been described rarely.

PCSK9 MONOCLONAL ANTIBODY

Mechanism of Action and Efficacy

PCSK9 is a resident protein in the cell membrane that binds to the LDL receptor and carries it to the lysosome for destruction, preventing it from recycling to the membrane. Naturally occurring loss of function mutations lead to very low levels of circulating LDL that are associated with significant reduction in morbidity and mortality from coronary vascular disease. Humanized monoclonal antibodies to PCSK9 inhibit its activity leading to major reductions in LDL levels in individuals with at least some LDL receptor activity. These agents have applications in individuals who have clinical cardiovascular disease or primary hypercholesterolemia with statin intolerance or whose LDL-C is not at goal. They can be given with other lipid-lowering agents. Preliminary results from placebo-controlled trials reveal a reduction in aggregate new coronary events.

Drug Dosage

Two preparations are available, alirocumab and evolocumab. Dose of the former is 75 or 150 mg SQ every 2 weeks; dose of the latter is 140 mg SQ every 2 weeks.

Side-Effects

Hypersensitivity reactions, injection site reactions, a transient flu-like syndrome, upper respiratory infection, increased liver enzymes, muscle pain, and other side-effects have been reported.

INHIBITION OF MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN (MTP)

Mechanism of Action and Efficacy

A requisite step in the secretion of VLDL by liver is the lipidation with triglyceride of Apo B-100. This is mediated by MTP and is inhibited by the agent, lomitapide. Reduction in the secretion of VLDL results in lower levels of its metabolic product, LDL. Use of this MTP inhibitor is currently limited to patients with homozygous LDL receptor deficiency. It can be used with a statin.

Drug Dosage

Dosage begins at 5 mg orally daily and is increased to tolerance or a maximum of 40 mg daily.

Side-Effects

The most significant adverse effect is hepatic steatosis that reverses spontaneously in some patients.

TABLE 19–3   **The primary hyperlipoproteinemias and their drug treatment.**

	Single Drug[a]	Drug Combination
Primary Chylomicronemia Chylomicrons (VLDL may be increased)	Niacin, fibrate[b] if VLDL increased	Niacin plus fibrate[b] if VLDL increased
Familial Hypertriglyceridemia VLDL and perhaps chylomicrons increased	Niacin, fibrate	Niacin plus fibrate
Familial Combined Hyperlipoproteinemia VLDL increased LDL increased VLDL, LDL increased	Niacin, fibrate, reductase inhibitor Reductase inhibitor, niacin, ezetimibe Reductase inhibitor, niacin, fibrate	Two or three of the single drugs
Familial Dysbetalipoproteinemia VLDL remnants, chylomicron remnants increased	Niacin, fibrate, reductase inhibitor	Fibrate plus niacin or niacin plus reductase inhibitor
Familial Hypercholesterolemia LDL increased	Reductase inhibitor, ezetimibe, niacin, resin	Two or three of the single drugs
Familial Ligand-Defective Apo B LDL increased	Reductase inhibitor, niacin, ezetimibe	Niacin plus reductase inhibitor or ezetimibe
Lp(a) Hyperlipoproteinemia Lp(a) increased	Niacin	

[a]Single-drug therapy should be tried before drug combinations are used.

[b]Fibric acid derivative.

APO B ANTISENSE OLIGONUCLEOTIDE

Mechanism of Action and Efficacy

Mipomersen is an oligonucleotide directed at the messenger RNA for Apo B-100. It lowers LDL-C levels by about 25% and would be expected to be effective in the absence of LDL receptor activity. Its use is limited to patients with homozygous LDL receptor deficiency.

Drug Dosage

The dose is 200 mg SQ once weekly.

Side-Effects

Elevated transaminases and hepatic steatosis have been reported. It is contraindicated in individuals with liver disease.

COMBINED DRUG THERAPY (TABLE 19–3)

Combinations of drugs may be useful (1) when LDL and VLDL levels are both elevated; (2) in cases of hypercholesterolemia in which significant increases of VLDL occur during treatment with a bile acid–binding resin; (3) where a complementary effect is required to normalize LDL levels; and (4) when a hyperlipidemia is accompanied by primary HDL deficiency or an elevated level of Lp(a).

Niacin with Other Agents

The combination of a fibrate with niacin may be more effective than either drug alone in managing marked hypertriglyceridemia.

Niacin usually normalizes triglycerides in individuals who have increased levels of VLDL while taking resins. The combination of niacin and resins is more effective than either agent alone in decreasing LDL levels in familial hypercholesterolemias. The combination is also very useful in the treatment of FCH. The absorption of niacin from the intestine is unimpeded by the resin. Niacin is complementary with ezetimibe in treating combined hyperlipidemia.

HMG-CoA Reductase Inhibitors with Other Agents

The addition of resin or niacin to a reductase inhibitor further decreases plasma levels of LDL in patients with primary hypercholesterolemias. Liver function and plasma CK activity should be monitored frequently when the combination includes niacin. These three drugs used together are more effective, frequently at lower doses, than any of their binary combinations in reducing LDL. Ezetimibe is synergistic with reductase inhibitors. The compatibility of rosuvastatin or pravastatin with fenofibrate makes these combinations useful in combined hyperlipidemia.

REFERENCES

Lipoproteins and Atherosclerosis

Clarke R, Peden JF, Hopewell JC, et al. PROCARDIS Consortium. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med.* 2009;361:2518. [PMID: 20032323]

Crosby J, Peloso GM, Auer PL, et al. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med.* 2014;371:22. [PMID: 24941081]

Guey LT, Pullinger CR, Ishida BY, et al. Relation of increased prebeta-1 high-density lipoprotein levels to risk of coronary heart disease. *Am J Cardiol*. 2011;108:360. [PMID: 21757044]

Khetarpal SA, Rader DJ. Triglyceride-rich lipoproteins and coronary artery disease risk. New insights from human genetics. *Arterioscler Thromb Vasc Biol*. 2015;35:e3. [PMID: 25573854]

Kohan AB. Apolipoprotein C-III: a potent modulator of hypertriglyceridemia and cardiovascular disease. *Curr Opin Endocrinol Diabetes Obes*. 2015;22:119. [PMID: 25692924]

Kruit JK, Brunham LR, Verchere CB, Hayden MR. HDL and LDL cholesterol significantly influence beta-cell function in type 2 diabetes mellitus. *Curr Opin Lipidol*. 2010;21:178. [PMID: 20463468]

Nduka C, Sarki A, Uthman O, Stranges S. Impact of antiretroviral therapy on serum lipoprotein levels and dyslipidemias: a systematic review and meta-analysis. *Int J Cardiol*. 2015;199:307. [PMID: 26241636]

Rohatgi A, Khera A, Berry JD, et al. HDL cholesterol efflux capacity and incident cardiovascular events. *N Engl J Med*. 2014;371:2383. [PMID: 25404125]

Vongpromek R, Bos S, Ten Kate GJ, et al. Lipoprotein(a) levels are associated with aortic valve calcification in asymptomatic patients with familial hypercholesterolaemia. *J Intern Med*. 2015;278:166. [PMID: 25487646]

Treatment

Burton BK, Balwani M, Feillet F, et al. A phase 3 trial of sebelipase alpha in lysosomal acid lipase deficiency. *N Engl J Med*. 2015;373:1010. [PMID: 26352813]

Cannon CP, Blazing MA, Giugliano RP, et al. IMPROVE-IT Investigators. Ezetimibe added to statin therapy after acute coronary syndromes. *N Engl J Med*. 2015;372:2387. [PMID: 26039521]

Chapman MJ, Stack JK, Ginsberg HN. PCSK9 inhibitors and cardiovascular disease: heralding a new therapeutic era. *Curr Opin Lipidol*. 2015;26:511. [PMID: 26780005]

Expert Dyslipidemia Panel of the International Atherosclerosis Society. An International Atherosclerosis position paper: global recommendations for the management of dyslipidemia. 2014;8:29. [PMID: 24528682]

Gaudet D, Alexander VJ, Baker BF, et al. Antisense inhibition of Apolipoprotein C-III in patients with hypertriglyceridemia. *N Engl J Med*. 2015;373:438. [PMID: 26222559]

Jacobson TA, Ito MK, Maki KC, et al. National Lipid Association recommendations for patient-centered management of dyslipidemia: Part 1-executive summary. *J Clin Lipidol*. 2014;8(5):473. [PMID: 25234560]

Jacobson TA, Maki KC, Orringer CE, et al. On behalf of the NLA Expert Panel. National Lipid Association recommendations for patient-centered management of dyslipidemia: Part 2. *J Clin Lipidol*. 2015;9:S1. [PMID: 26699442]

Rader DJ, Kastelein JI. Lomitapide and mipomersen: two first-in-class drugs for reducing low-density lipoprotein cholesterol in patients with homozygous familial hypercholesterolemia. *Circulation*. 2014;129:1022. [PMID: 24589695]

Reith C, Armitage J. Management of residual risk after statin therapy. *Atherosclerosis*. 2016;245:161. [PMID: 26722833]

Richardson K, Schoen M, French B, et al. Statins and cognitive function: a systematic review. *Ann Intern Med*. 2013;159:688. [PMID: 24247674]

Robinson JG, Farnier M, Krempf M, et al. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *N Engl J Med*. 2015;372:1489. [PMID: 25773378]

Stein EA, Giugliano RP, Koran MJ, et al. Efficacy and safety of evolucumab (ANG 145), a fully human monoclonal antibody to PCSK9 in hyperlipidemic patients on various background lipid therapies: pooled analyses of 1359 patients in four phase 2 trials. *Eur Heart J*. 2014;35:2249. [PMID: 24598985]

Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guidelines on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults. A report of the American College of Cardiology/American Heart Association task force on practice guidelines. Circulation. 2014;129(25 supplement 2):S1-45. Erratem: *Circulation*. 2015;132(25):E396. [PMID: 24222016]

This page intentionally left blank

Obesity

Alka M. Kanaya, MD and Christian Vaisse, MD, PhD

α-MSH	α-Melanocyte–stimulating hormone		**NHLBI**	National Heart, Lung, and Blood Institute
ACTH	Adrenocorticotropic hormone		**NTRK2**	Neurotrophic tyrosine kinase, receptor, type 2
AGRP	Agouti-related peptide		**NPY**	Neuropeptide Y
ARC	Arcuate nucleus of the hypothalamus		**OSA**	Obstructive sleep apnea
BDNF	Brain-derived neurotrophic factor		**PAI-1**	Plasminogen-activated inhibitor
BMI	Body mass index		**PC-1**	Proconvertase 1
CCK	Cholecystokinin		**PCOS**	Polycystic ovarian syndrome
CCK-1R	CCK 1 receptor		**POMC**	Pro-opiomelanocortin
CSF	Colony-stimulating factor		**PVN**	Paraventricular nucleus
FFA	Free fatty acid		**PYY**	Peptide YY
FTO	Fat mass- and obesity-associated gene		**RYGB**	Roux-en-Y gastric bypass
GI	Gastrointestinal		**SH2B1**	Adapter protein 1
GLP1	Glucagon-like peptide 1		**SOS**	Swedish Obese Subjects
IL-6	Interleukin 6		**SIM1**	Single minded 1
LepR	Leptin receptor		**STAT**	Signal transducer and activator of transcription
LHA	Lateral hypothalamic area		**T2DM**	Type 2 diabetes mellitus
MC4R	Melanocortin-4 receptor		**TMEM18**	Transmembrane protein 18
MCP-1	Monocyte-chemoattractant protein-1		**TNFα**	Tumor necrosis factor-α
NAFLD	Nonalcoholic fatty liver disease		**WHO**	World Health Organization
NCEP	National Cholesterol Education Program			
NEGR1	Neuronal growth regulator 1			
NHANES	National Health and Nutrition Examination Survey			

DEFINITION AND EPIDEMIOLOGY

Definition

Obesity and overweight categories have been defined by examining longitudinal study data that associate a given weight with future adverse health effects. The currently accepted surrogate measure of body fatness is body mass index (BMI) which is measured as weight in kilograms divided by height in meters squared. For adults, a BMI less than 18.5 kg/m^2 is underweight, 18.5 to 24.9 kg/m^2 is healthy weight, 25.0 to 29.9 kg/m^2 is overweight, and greater than or equal to 30 kg/m^2 is obese. For children and adolescents, a BMI between the 85th and 95th percentile for age

and sex is considered at risk of overweight, and BMI at or above the 95th percentile is considered overweight or obese.

However, BMI does not account for ethnic differences in skeletal structure or musculature. Body frame size varies dramatically by race/ethnicity from small-framed East Asian adults to larger framed Pacific Islanders. Moreover, conventional cut-points for adult overweight (BMI ≥25 kg/m^2) and obesity (BMI ≥30 kg/m^2) do not correspond to similar absolute or relative metabolic risk in all ethnic group. As a result, the World Health Organization (WHO) and International Obesity Task Force proposed a lower BMI cut-point to define obesity in South and East Asian adults: 23 kg/m^2 for overweight and 25 kg/m^2 for obesity.

Prevalence and Projections

National surveys performed by the Centers for Disease Control and Prevention in the United States have found a significant increase in the prevalence of overweight and obesity over the past 30 years. In 2003 to 2004, according to data from the National Health and Nutrition Examination Survey (NHANES), 32.2% of adults were obese. Trends among adults have been tracked by repeat NHANES surveys, showing that the percent of overweight or obese has climbed from 47.0% in 1976 to 1980 to 66.2% in 2003 to 2004, and the prevalence of obesity alone has more than doubled from 15.0% to 32.9% overall. In 2015, the prevalence of obesity was 48.1% among non-Hispanic blacks, 42.5% among Hispanics, and 34.5% among non-Hispanic whites.

The trends in obesity have been most striking among children and adolescents with two- to threefold increases in rates over the past three decades in all age groups. In the NHANES 2003 to 2004 survey, a total of 17.1% of U.S. children and adolescents were overweight (BMI ≥95th percentile). However, the newest trend reports by the Centers for Disease Control show that no change in obesity prevalence has been observed between the 2011-2012 and 2013-2014 data collection period.

While the United States leads most of the world with high prevalence of overweight and obesity, these conditions are becoming more prevalent worldwide, even in developing countries. In China, overweight prevalence in adults increased from 14.6% to 21.8% between 1992 and 2002, with greatest increases seen among men, urban residents, and high-income groups. In India, the prevalence of overweight ranged from 9.4% in rural men to 38.8% in urban women in 2003 to 2005.

Globally, according to a pooled analysis, in 2005 there were an estimated 937 million (95% CI, 922-951 million) overweight adults and 396 million (388-405 million) obese adults. Using conservative projections, by 2030 the number of overweight adults was estimated to be 1.35 billion and 573 million obese individuals without adjusting for secular trends. The authors warned that if recent secular trends continue unabated, the absolute numbers may total 2.16 billion overweight and 1.12 billion obese adult individuals. Among U.S. residents, projections forecast that by 2030, 86.3% adults will be overweight or obese (51.1%) with black women (96.9%) and Mexican American men (91.1%) being most affected. Wang et al estimate that by 2048, all American adults would become overweight or obese, and in children, the prevalence of overweight (BMI ≥95th percentile) would almost double to 30%.

Possible Explanations for the Increased Obesity Rates

Several hypotheses have been proposed to explain the rapid rise in overweight/obesity in developed and developing countries. The simple explanations start with changes in lifestyle behaviors, primarily the lack of regular physical activity and increased caloric intake that tip the energy balance equilibrium. The reality of the factors promoting the obesity epidemic is clearly more complex and involves multiple layers of determinants including social norms and values, several sectors of influence from governmental to media to the food/entertainment industries, behavioral settings including the work/home/neighborhood environment, to individual factors that include genetic/epigenetic, microbiomic, socioeconomic, cultural, and psychosocial factors. In a novel social network analysis of the Framingham cohort, investigators found that weight status (gain, loss, or stability of weight) for an individual was highly influenced by his/her friends and family extending out to three degrees of separation. A multifaceted approach will be required to meaningfully curb the obesity epidemic globally.

PATHOPHYSIOLOGY AND GENETICS OF OBESITY

Regulation of Food Intake and Energy Expenditure

Obesity is an increase of energy stored as fat that occurs when caloric intake exceeds caloric expenditure. What causes this imbalance is less clear, but recent advances in our understanding of the physiological systems responsible for the maintenance of energy stores in response to variable access to nutrition and demands for energy expenditure have provided some insights into the pathophysiology of obesity. The physiological system controlling food intake and energy expenditure is composed of (1) long-term and short-term afferent signals that allow for sensing the energy status of the individual; (2) integrating brain centers, most importantly within the hypothalamus, where the level of the efferent response is determined; and (3) efferent signals including those regulating the intensity of hunger and the level of energy expenditure.

One common misconception is that this physiological system is dedicated to the prevention of obesity. Instead this system's essential role is in the prevention of starvation (ie, ensuring adequate energy intake to compensate for the energy requirements of basal metabolism, physical activity, growth, and reproduction). As a result, this physiological system is more strongly biased toward prevention of energy deficiency rather than excess storage.

Informing the Brain of the Energy Status: Leptin and Short-Term Gastrointestinal Signals

Leptin   The major afferent signal allowing the brain to sense the level of energy stores is the hormone leptin. Discovered in 1994, this cytokine-like 167 amino acid protein is released by adipocytes. Under basal conditions the circulating levels of serum leptin are correlated with fat mass (r = 0.8) and decrease after weight loss. Decreasing leptin levels inform the brain of diminishing fat storage resulting from negative energy balance. This results in compensatory effects on appetite and energy expenditure, aimed at replenishing the stores and reestablishing energy balance. The role of leptin is best understood from the phenotype of rare cases of humans with complete leptin deficiency. Despite early-onset, extremely severe obesity, these patients show behaviors and

physiological signs of starvation. This includes hyperphagia, decreased immune function, and hypogonadotropic hypogonadism. The latter reflects deficiency of the leptin signal informing the brain that sufficient energy stores are available for reproduction. Hormonal replacement of leptin-deficient patients corrects all these abnormalities, abolishing the hyperphagia leading to normalization of weight and maturation of the reproductive axis.

Leptin's action is mediated by the leptin receptor (LepR). This receptor is a member of the class I cytokine receptor family which also includes the growth hormone, prolactin and erythropoietin receptors (see Chapter 1). These receptors have a single transmembrane domain and homodimerize upon binding of the ligand. Leptin binding to LepR leads to phosphorylation of the JAK2 receptor-associated kinase followed by recruitment and phosphorylation of the signal transducer and activator of transcription STAT3. Upon activation STAT3 dimerizes and translocates to the nucleus, where it activates transcription of target genes. The LepR is expressed in almost every tissue but only one of five isoforms, LepRb, contains the STAT3 recruitment domain. This isoform is expressed specifically in leptin-responsive brain regions.

In summary, leptin functions as a long-term signal of energy balance by informing the brain of changes in the level of energy stored as fat. A perceived decrease in leptin levels increases the amount of food consumed and minimizes energy expenditure. Leptin levels do not change with meals, and leptin does not acutely change meal size.

Signals from the gastrointestinal tract In addition to leptin, which represents a long-term adiposity signal, a number of other hormones, most of which are secreted by the gastrointestinal (GI) system, regulate appetite on a short-term basis (ie, in response to a meal) (Table 20–1). These hormones act on **satiation** (the feeling of fullness that contributes to the decision to stop eating) rather than on **satiety** (the prolongation of the interval until hunger or a drive to eat reappears). Some of these hormones have been shown to act on the same brain centers as leptin does. Altered secretion of or response to these GI hormones may directly contribute to the pathogenesis or the maintenance of obesity. Cholecystokinin (CCK) is secreted by duodenal I cells into the bloodstream. Its secretion is stimulated by the presence of fat and protein in the duodenum. CCK stimulates gut motility, contraction of the gallbladder, pancreatic enzyme secretion, gastric

emptying, and acid secretion. CCK acts locally, stimulating vagal sensory nerves through the CCK 1 receptor (CCK-1R) and informs the brain that fat and proteins are being processed and will soon be absorbed. This message is conveyed to the hypothalamus via the hindbrain. Circulating CCK levels increase after meals and infusion of a CCK-1R agonist (CCK33), to postprandial levels, suppresses food intake. Conversely, infusion of a CCK-1R antagonist increases calorie intake. CCK is also a proximal mediator of the broader satiation process, attenuating meal-induced changes of more distal GI hormones, such as ghrelin and peptide YY (PYY).

Glucagon-like peptide 1 (GLP1) is a 30 amino acid peptide derived from the proglucagon gene. GLP1 is secreted into the bloodstream by the intestinal L-cells of the ileum and colon in response to the presence of nutrients in the lumen. GLP1 has a very short half-life (2 minutes), and is degraded by dipeptidyl peptidase-4. It increases insulin secretion in a glucose-dependent manner, decreases glucagon secretion, increases beta cell mass, inhibits gastric emptying, and decreases food intake. GLP1 is thought to signal through the hindbrain via stimulation of the GLP1 receptor (GLP1R) on the vagus nerve.

PYY, a 36 amino acid peptide is released into the bloodstream by intestinal endocrine L-cells of the distal gut (ileum and colon) after food ingestion and in response to the presence of fat in the lumen. PYY3-36, the major form of circulating PYY, binds to the hypothalamic NPY-Y2 receptor, and reduces food intake. Obese individuals have lower PYY levels after a test meal than nonobese controls.

Ghrelin, a 28 amino acid octanoylated peptide, is secreted into the bloodstream by endocrine cells lining the fundus of the stomach. Ghrelin secretion is stimulated by fasting, increases preprandially, and is suppressed by food intake. Ghrelin is orexigenic, increasing food intake when administered peripherally, and acts in part by directly modulating the activity of the NPY (neuropeptide Y)/ AGRP (agouti-related peptide) neurons in the arcuate nucleus of the hypothalamus.

Central Integration of Energy Homeostasis Signals

The major challenge of adapting food intake to energy expenditure lies in the discontinuous availability of the first and the continuous

TABLE 20–1 Selected GI and pancreatic peptides that regulate food intake.

	Main Site of Synthesis	Secretion Pattern	Effect on Food Intake
Ghrelin	Gastric X/A-like cells	Increases prior to meal; decreased by food intake	Stimulates food intake
Cholecystokinin (CCK)	Proximal intestinal I cells	Stimulated by duodenal presence of fat and protein	Promotes meal termination and reduces meal size
Peptide YY3-36 (PYY)	Distal-intestinal L cells	Stimulated by presence of fat in the lumen	Reduces appetite and food intake
Glucagon-like peptide 1 (GLP-1)	Distal-intestinal L cells	Stimulated by presence of nutrients in the lumen	Short-term inhibition of food intake
Pancreatic polypeptide (PP)	Pancreatic F cells	Released in proportion to calories ingested	Reduces appetite and food intake

and variable levels of the second required for survival, growth, and reproduction. The current prevailing model is that the concentration of circulating adiposity signals, primarily leptin, influences the response to short-term satiation signals thereby allowing for sufficient energy intake to maintain a constant level of stored energy. Genetic studies in mice and humans have outlined several critical neuronal populations and circuits that translate information provided by afferent circulating hormonal signals into neural responses that regulate appetite and energy expenditure. The primary locus for the integration of peripheral signals that influence energy balance is the hypothalamus (Figure 20–1). Specifically, two adjacent groups of neurons in the arcuate nucleus (ARC) of the hypothalamus, characterized by their production of specific neuromediators, act as the primary site in the brain for receiving the humoral signals that reflect body energy status. The AGRP/NPY neurons are orexigenic. They are inhibited by leptin and their activity is also modulated by GI hormones, such as ghrelin and PYY. The proopiomelanocortin (POMC) neurons are anorexigenic and are activated by leptin. In these neurons, POMC is processed by the specific proteases proconvertase-1 and proconvertase-2 to the anorexigenic neuropeptide α-melanocyte–stimulating hormone (α-MSH). Interactions between these neurons within the ARC allows for cross talk and modulation of the neuronal output.

These first-order neurons send projections to second-order neurons in other regions of the hypothalamus, specifically the paraventricular nucleus (PVN) and the lateral hypothalamic area (LHA), as well as to the hindbrain. Both α-MSH and AGRP act on a common receptor, the melanocortin-4 receptor (MC4R) expressed on PVN neurons. α-MSH binds to and activates the melanocortin 4 receptor (MC4R) and by competing with α-MSH for the MC4R, AGRP inhibits MC4R-bearing neurons. The MC4R neurons then innervate neurons in other areas of the hypothalamus and brain stem to produce an integrated inhibition of appetite and increase in energy expenditure.

Leptin Resistance in Obesity

While recombinant leptin can correct the exceptionally rare cases of obesity due to genetic leptin deficiency, all other obese patients have leptin levels that are proportional to their fat mass. In addition, recombinant leptin injections in obese patients do not lead to weight loss. Therefore, despite the presence of elevated leptin concentrations, which should reduce food intake and body fat, obese patients appear to be insensitive or resistant to leptin. Both genetic and environmental factors have been shown to contribute to this leptin resistance, and a number of specific alterations in the downstream leptin effector pathways have been suggested to participate in this obesity-associated leptin resistance. Understanding the nature of this leptin resistance may allow for the development of novel obesity treatments, but the molecular basis of obesity is probably heterogenous and different in each patient.

Genetics of Obesity

Excess caloric intake and the tendency toward a more sedentary lifestyle are certainly to blame for the increased prevalence
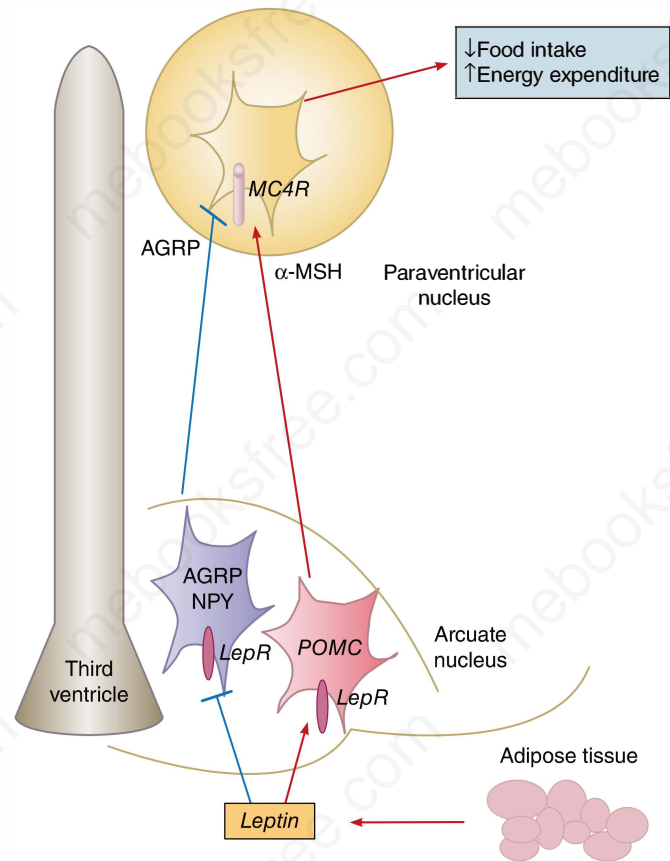
FIGURE 20–1 Leptin-melanocortin system of energy balance. Information about the body's energy stores is conveyed to the brain by hormones such as leptin. Leptin is secreted by adipocytes in proportion to the body's fat mass. Leptin binds to its receptors on two populations of neurons in the arcuate nucleus of the hypothalamus: the orexigenic AGRP/NPY–expressing neurons, and the anorexigenic POMC-expressing neurons. These two groups of neurons have projections to the paraventricular nucleus of the hypothalamus, as well as to other regions of the brain. The PVN has a dense neuronal population that expresses MC4R. When leptin binds its receptors on POMC neurons, α-MSH, a cleavage product of POMC, is released. Activation of MC4R in the PVN by α-MSH relays a satiety signal and causes a decrease in food intake. AGRP is an antagonist of MC4R, and competes with α-MSH for binding to MC4R. Binding of AGRP to MC4R leads to increased food intake. Leptin activates POMC neurons and inhibits AGRP neurons. Therefore, by activating its receptors on these two neuronal populations, leptin acts in a concerted way to increase MC4R activation by α-MSH and decrease its antagonism by AGRP resulting in a net decrease in food intake. (α-MSH, α-melanocyte–stimulating hormone; AGRP, agouti-related peptide; ARC, arcuate nucleus; MC4R, melanocortin 4 receptor; NPY, neuropeptide Y; POMC, pro-opiomelanocortin; PVN, paraventricular nucleus; LepR, leptin receptor.)

of obesity, but individuals exposed to the same environmental pressures have different levels of vulnerability. Indeed, genetic epidemiologic studies, such as twin studies and adoption studies have implicated genetic factors in the susceptibility to obesity. These studies have shown that genetic factors account for 40% to

TABLE 20–2 Common variants increasing the risk of obesity.[a]

Chromosomal Localization	Nearby Genes	Odds Ratio of Obesity in Adults
16q12	FTO	1.25
18q21	MC4R	1.14
2	TMEM18	1.20
16p11	SH2B1	1.10
1p31	NEGR1	1.06
11p14	BDNF	1.05

[a]Only variants replicated in multiple studies and with odds ratios of obesity above 1.05 are indicated.

70% of the population variation in BMI and that the heritability of obesity increases with its severity. Emphasis has, therefore, now shifted from the question of whether human obesity has a genetic component to how many and which genetic variants underlie this susceptibility. While the majority of genes that contribute to the predisposition to obesity are still unknown, both common genetic variants, with small effects on the BMI at the level of the population, and single gene defects, causing Mendelian forms of obesity, have been described.

Large and comprehensive genome-wide association studies find that a number of common genetic variants are associated with obesity (Table 20–2). These variants are near genes that are expressed in the central nervous system, further suggesting a role for an alteration of central regulatory pathways in the pathogenesis of obesity. However, since these variants are not located within the protein-coding regions of these genes, the exact mechanism by which they predispose to obesity remains to be elucidated. More importantly, the individual as well as the combined effects of these common variants only explain a small fraction of the inherited variability in obesity (10%-20%), suggesting that rare variants may contribute more significantly to the genetic predisposition to this condition.

Indeed a number of single gene defects that cause human obesity have been uncovered. These defects have provided more insight into the genetic predisposition and pathophysiology of obesity. Certain mutations can cause obesity as part of a complex syndrome, in which developmental phenotypes are prominent (Table 20–3). However, most single gene defects causing Mendelian forms of obesity are in genes of the leptin-melanocortin system (Table 20–4). The description of patients with these mutations has both provided insight into the genetic predisposition to obesity and has confirmed that the hypothalamic leptin-melanocortin system is critical for energy balance in humans. Mutations in leptin and the LepR cause severe obesity associated with hypogonadotropic hypogonadism. Mutations in POMC cause severe obesity, due to absent synthesis of α-MSH in the hypothalamus, and neonatal adrenal insufficiency, due to lack of ACTH production in the anterior pituitary gland. These recessive syndromes are extremely rare. In contrast, mutations in MC4R represent the most common genetic cause of severe obesity accounting for 2.5% of all cases. Patients who carry such mutations do not have specific clinical or biological characteristics that differentiate them from other patients with severe obesity; genetic testing is the only reliable method to identify these patients. At present, such testing is not recommended since the presence or absence of one of these mutations carries no specific implications for the clinical management of these patients.

HEALTH CONSEQUENCES OF OBESITY

Mechanism Underlying Obesity Complications: Adipose Tissue as an Endocrine Organ

The massive public health burden of obesity is due to the psychological, behavioral, and medical consequences of this chronic increase in adipose tissue mass. Pathophysiological mechanisms underlying the complications of obesity include the psychological and behavioral responses to the image projected in a society in

TABLE 20–3 Syndromes that cause human obesity.

Name/OMIM	Clinical Features	Prevalence	Genetics
Prader-Willi OMIM#176270	Diminished fetal activity, obesity, muscular hypotonia, mental retardation, short stature, hypogonadotropic hypogonadism, small hands, and feet	1/25,000	Imprinting + Deletion 15q12
Bardet-Biedl OMIM#209900	Hypogonadism, pigmentary retinopathy, polydactyly, mental retardation, obesity	1/100,000	Recessive >12 genes involved in function of the primary cilium
Alström OMIM#203800	Myocardiopathy, sensory deficit (retinopathy, deafness), obesity dyslipidemia, diabetes	Rare	Recessive ALMS1, 2p13
Börjeson-Forssman-Lehman OMIM#301900	Severe mental defect, epilepsy, morbid obesity, hypogonadism, facial dysmorphy	Very rare	X-linked recessive PHF6, Xq26.3
Cohen OMIM#216550	Mental retardation, characteristic facies, obesity	Very rare	Recessive COH1, 8q22
Carpenter OMIM#201000	Acrocephalosyndactyly, polydactyly, obesity, mental retardation	Very rare	Recessive RAB23, 6p11

TABLE 20–4 Monogenic obesity due to mutations in genes related to the leptin-melanocortin system.

Gene	N = Families	Transmission	Obesity	Other Phenotypes
Leptin	5(2)	Recessive	Very severe, early onset	Hypogonadotropic hypogonadism
Leptin receptor	3(1)	Recessive	Very severe, early onset	Hypogonadotropic hypogonadism
POMC	5(5)	Recessive	Very severe, early onset	ACTH deficiency
PC-1	2(2)	Recessive	Severe, early onset	Hyperproinsulinemia, hypercortisolism, hypogonadotropic hypogonadism
NTRK2	1	Dominant	Very severe, early onset	Mental retardation
SIM1	3(3)	Dominant	Very severe, early onset	+/– Mental retardation
MC4R	2.5% of severely obese patients	Dominant	Severe, variable onset	None

which obesity is considered cosmetically unattractive, the mechanical effects of increased adipose mass on weight-bearing joints as well as other organs, and the metabolic consequences of increased circulating free fatty acids (FFAs).

It is also now clear that the adipose cell is not simply dedicated to inertly storing energy as triglycerides and releasing free fatty acids and glycerol during periods of increased energy needs. In addition to leptin, described earlier, which informs the brain of the level of energy stores, adipose cells secrete a number of factors, known as adipokines, which can play a role in the pathophysiology of obesity and its complications (Figure 20–2). For example, adiponectin is an abundant plasma protein, structurally related to the complement 1q family, which is secreted exclusively from adipose tissue. Adiponectin circulates in serum as a range of multimers from low-molecular-weight trimers to high-molecular-weight (HMW) dodecamers. Circulating levels of adiponectin are negatively correlated with fat mass and are reduced in obesity. Adiponectin improves whole-body insulin sensitivity through its receptor-mediated actions on the muscle and the liver. Specifically, adiponectin stimulates fatty acid oxidation and glucose uptake in skeletal muscle and suppresses glucose output in the liver.

In addition, a number of factors secreted by adipose tissue, particularly in the context of obesity, are proinflammatory cytokines. It is now generally accepted that obesity is a state of chronic, low-grade inflammation. These inflammatory molecules exert their effects through both paracrine and endocrine pathways. Initially thought to be secreted by adipocytes, most of these factors are actually produced by other cells such as macrophages, which infiltrate the adipose tissue in obesity. Tumor necrosis factor-α (TNF-α) was the first adipose-derived factor suggested to represent a link between adiposity and the insulin resistance associated with type 2 diabetes. Released by activated macrophages, which infiltrate expanding adipose tissue in obesity, this cytokine has been strongly implicated in the pathogenesis of insulin resistance, directly impairing insulin signaling in hepatocytes and adipocytes. Other adipokines involved in obesity-associated inflammation include interleukin 6 (IL-6), monocyte-chemoattractant protein-1 (MCP-1), plasminogen-activated inhibitor (PAI-1), and colony-stimulating factor (CSF).

Metabolic Complications of Obesity: Insulin Resistance and Type 2 Diabetes

Hyperinsulinemia is characteristically associated with obesity. This hyperinsulinemia is linked to insulin resistance (ie, the decrease in insulin-stimulated glucose utilization, which increases proportionally with fat mass). One of the major consequences of this obesity-associated insulin resistance is the requirement for increased insulin secretion. While the majority of obese individuals can secrete sufficient insulin to accommodate the insulin-resistant state, this increased burden on the insulin-producing pancreatic beta cells increases the susceptibility to failure, resulting in type 2 diabetes (T2DM) in genetically predisposed individuals. Indeed, over 80% of T2DM patients are obese. In comparison with a normal-weight individual, BMI greater than 40 (obesity class 3) in a person under 55 years of age increases the risk of T2DM by 18.1-fold in men and 12.9-fold in women. Classification as overweight (BMI 25-30) increases the prevalence of T2DM three- to fourfold in both men and women under 55 years of age, and BMI greater than or equal to 25 kg/m^2 has been the recommended criterion for diabetes screening. However, in 2015 the American Diabetes Association revised this criterion for diabetes screening in Asians to be BMI greater than or equal to 23 kg/m^2, based on several studies showing higher diabetes risk at lower BMI in most Asian groups. The relative risks associated with excess body weight

Leptin
• Appetite & energy balance
• Reproduction
• Insulin sensitivity
• Immunity

PAI-1
Hemostasis

Adiponectin
Insulin sensitivity

Angiotensinogen
Blood pressure

Pro-inflammatory cytokines
Insulin sensitivity
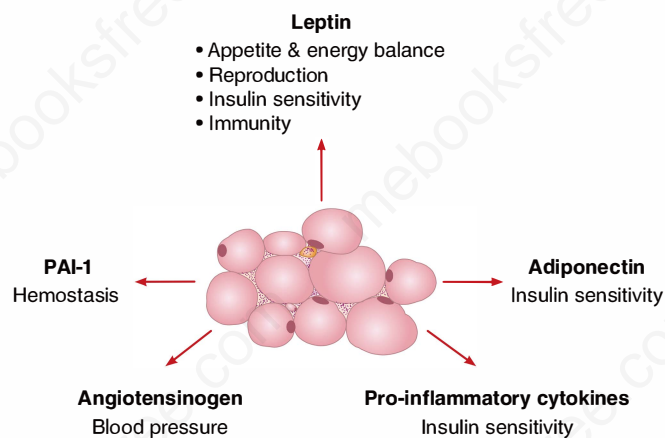
FIGURE 20–2 Adipose tissue secretions influencing metabolism.

are weaker in older individuals but are still present (eg, about twofold for overweight and three- to sixfold for obesity class 3).

The increased rate of T2DM is the most worrisome public health issue linked to the increased rates of obesity. The anticipated health outcomes such as premature heart disease, peripheral vascular disease, renal failure, blindness, are deeply concerning and place this among our highest contemporary public health priorities. Some recent studies have estimated that the current generation of American youth will be the first to exhibit shorter lifespans than their parents, because of the impact of obesity on T2DM.

Dyslipidemia

Obesity is associated with dyslipidemia characterized by increased levels of very low density lipoprotein (VLDL) cholesterol, triglycerides, and total cholesterol as well as a decrease in HDL cholesterol and an increase in small dense LDL particles. The difference in serum triglyceride concentrations between individuals with BMI less than 21 and those with BMI greater than 30 is about 65 mg/dL (in women) and 62 to 118 mg/dL (in men, depending on age). For HDL cholesterol, each 1-unit BMI change is associated with an HDL-cholesterol decrease of 1.1 mg/dL for young men and 0.7 mg/dL for young women. All of the components of this obesity-associated dyslipidemia have been shown to be atherogenic and play a major role in the development of atherosclerosis and cardiovascular disease in obese individuals.

Weight loss and exercise, even if they do not result in normalization of body weight, can improve this dyslipidemia and thus reduce CVD risk. In addition, obese individuals should be targeted for intense lipid-lowering therapy, when necessary.

The Metabolic Syndrome

The concept of the **metabolic syndrome** was first introduced in 1923 by Kylin who described the clustering of hypertension, hyperglycemia, and gout as a syndrome. Reaven and others redefined this concept as the insulin resistance syndrome, a clustering of glucose intolerance, high triglycerides, low HDL cholesterol, and hypertension. The two most consistent characteristics of persons with the syndrome are insulin resistance and central abdominal obesity. There is strong epidemiologic and pathophysiologic evidence linking visceral adiposity with insulin resistance and diabetes, with dyslipidemia, hyperinsulinemia, and decreased fibrinolysis as possible mediators of the relationship between visceral fat and type 2 diabetes and cardiovascular disease risk.

The earliest attempt to create a definition for this constellation of metabolic disorders was by the WHO wherein an individual with glucose intolerance or insulin resistance was required to have two or more criteria (elevated triglycerides, low HDL cholesterol, high BMI or waist-hip ratio, hypertension, or microalbuminuria) in order to be categorized as having metabolic syndrome. The next attempt at a clinical definition was by the National Cholesterol Education Program (NCEP)-Adult Treatment Panel III in 2001 that required three or more of the following criteria for the metabolic syndrome definition: high fasting glucose, high triglycerides, low HDL, high blood pressure, or waist circumference.

The International Diabetes Federation (IDF) proposed further refined criteria that require central abdominal obesity, and two more criteria as proposed by the NCEP. The IDF criteria were the first ones to account for the heterogeneity of body size among different ethnic groups, and the committee recommended different waist circumference cut-points based on ethnicity with a revised definition for South Asians, Chinese, and Japanese. They recommended similar waist circumference cut-points in men (≥90 cm) and in women (≥80 cm) among these various ethnic groups, compared to the European cut-points of greater than or equal to 94 cm in men and greater than or equal to 80 cm in women. More recently, the utility of the metabolic syndrome has been questioned, given that there are no treatment options available, and rather each cardiometabolic risk factor (eg, blood pressure control) is treated individually.

Cardiovascular Complications

Higher cardiovascular disease risk has been linked to visceral and hepatic adiposity with several distinct mechanisms including inflammation, thrombosis, hyperglycemia, atherogenic dyslipidemia, and adipocytokines (hormones and proteins exclusively secreted by adipose tissue). Obesity has been implicated as one of the major risk factors for hypertension, heart failure, and coronary heart disease. Other cardiovascular conditions associated with obesity include peripheral arterial disease and atrial fibrillation.

Altered hemodynamics of obesity due to increases in total blood volume, systemic arterial pressure, and cardiac output increase the workload of the heart and can lead to adverse effects on cardiovascular structure and function. Overweight individuals develop left ventricular dilation and hypertrophy and left atrial enlargement. These structural abnormalities increase the risk of heart failure and of atrial fibrillation. In the Framingham Heart Study, every 1-unit increase in BMI increased the risk of heart failure by 5% in men and 7% in women. However, several recent studies have also found an obesity-cardiovascular disease paradox where overweight and obese individuals with existing heart failure or heart disease have a lower risk of mortality than those who are underweight or normal weight. These findings suggest that obese individuals may have the metabolic reserve to withstand high catabolic needs as exist in heart failure.

Pulmonary Complications

Obesity can cause alveolar hypoventilation, increased pulmonary blood volume, and mechanical effects on the respiratory system. Obstructive sleep apnea (OSA) is associated with obesity and other features of the metabolic syndrome, with the most closely correlated factor being hypertension. Intermittent upper airway obstruction results in hypoxia, sympathetic nervous system surges, and airway inflammation. Patients with sleep apnea have higher risk of systemic and pulmonary hypertension, heart failure, dysrhythmias, myocardial infarction, and overall mortality. Mechanisms that have been implicated in the development of cardiovascular disease in OSA include increased oxidative stress and inflammation that cause endothelial dysfunction. While significant weight loss can improve OSA and associated metabolic

features, mechanical therapies used to treat sleep apnea with constant positive airway pressure have not been found to improve hypertension or other metabolic features.

Gastrointestinal Complications

Nonalcoholic fatty liver disease (NAFLD) is commonly seen with insulin resistance or obesity and includes a spectrum of liver disease from steatosis to nonalcoholic steatohepatitis, advanced fibrosis, and rarely, cirrhosis. Fat accumulates within the liver with recruitment of inflammatory cells and fibrosis. The prevalence of NAFLD in obesity is 50% to 85%. Risk factors associated with NAFLD besides obesity include advancing age, visceral adiposity, atherogenic dyslipidemia, and hypertension. NAFLD is the leading cause of chronic liver disease in the United States, and given the high rates of obesity, approximately one-third of U.S. adults are estimated to have NAFLD.

NAFLD has been observed in overweight children and adolescents with prevalence estimated at 6% to 23% in population-based U.S. samples. In children, there seems to be a clear male predominance for NAFLD which suggests distinct pathophysiologic mechanisms for NAFLD between adults and children. Histologic criteria have been proposed to distinguish adult (type 1) and pediatric (type 2) NAFLD. In a large study of German children, the three main factors that contributed to NAFLD by principal components analysis included an insulin resistance/visceral adiposity factor, a body fat distribution/CRP factor, and lastly a sex steroid hormone factor.

Other GI complications of abdominal obesity include gastroesophageal reflux disease and hiatal hernia. Both of these conditions can cause erosive mucosal changes in the lower esophagus leading to higher risk of Barrett esophagus.

Reproduction and Gynecologic Complications

Obesity has been associated with both female and male infertility. In women, polycystic ovarian syndrome (PCOS) is the main cause of female infertility associated with obesity. Women with PCOS have ovulatory dysfunction and hyperandrogenemia. PCOS is reviewed in more detail in Chapter 13. In men, obesity can cause reduced sperm counts and motility. Hormonal changes with obesity and associated metabolic disorders may lead to poor sperm production. Obesity is also associated with erectile dysfunction through proinflammatory mechanisms and decreased testosterone levels.

Cancer

Several cancers have been linked to obesity, including breast, endometrial, colorectal, prostate, and renal cell carcinomas. Putative mechanisms for breast and endometrial cancer include higher circulating levels of unopposed estrogen and ovarian hyperandrogenism, causing higher levels of testosterone and lower levels of luteinizing hormone. The relationship between obesity and colon cancer is stronger in men than women.

MANAGEMENT OF THE OBESE PATIENT

Screening and Prevention of Complications

Periodic health screening with a measurement of both height and weight to calculate BMI is recommended by many organizations including the National Heart, Lung, and Blood Institute (NHLBI), the U.S. Preventive Services Task Force, and the WHO. Furthermore, the NHLBI also recommends waist circumference measurement routinely to find persons at risk of metabolic complications of overweight. Patients who are found to be in the overweight or obese categories of BMI are recommended to have other risk factors checked as well, including blood pressure, fasting glucose, and lipoprotein levels.

Therapeutic Approaches for Weight Loss

Several behavioral, pharmacologic, and surgical therapies for weight loss have been studied in randomized controlled trials. In general, behavioral interventions with lifestyle change including diet and exercise and pharmacologic therapies work to achieve modest amounts of weight loss while patients adhere to the intervention. Based on these studies, in 2012 the U.S. Preventive Services Task Force recommended that adults with BMI greater than or equal to 30 kg/m^2 be offered or referred for counseling and behavioral interventions to promote weight loss. Recommended for more severe levels of obesity, bariatric surgery does lead to larger amounts of weight loss, but requires lifelong medical supervision and management of the side effects associated with the surgery.

A. Behavior, diet, and exercise therapies   While many believe that long-term weight loss with lifestyle change is not achievable by most patients, longitudinal studies have found that 20% of overweight individuals are successful at weight loss defined as losing at least 10% of initial body weight and maintaining the loss for at least 1 year. Individuals enrolled in the National Weight Control Registry have lost an average of 33 kg and maintained the loss for more than 5 years. Behaviors that have been most effective for long-term weight loss include high levels of physical activity (~1 hour/d), self-monitoring weight, eating a low-calorie, low-fat diet, eating breakfast regularly, and maintaining a consistent eating pattern throughout the week. Wearable activity trackers and smart phone applications are useful aids for self-monitoring physical activity and dietary intake.

Dietary interventions   Several commercial, fad, and therapeutic dietary interventions exist for weight loss. Recent dietary trends have focused on relative changes in macronutrient composition of the diet to achieve weight loss, such as low-fat or low-carbohydrate diets. Few have been rigorously tested, and no trials have long-term outcome data (defined as the occurrence of comorbidities or mortality). A recent randomized trial comparing four popular diets (Atkins, Zone, Weight Watchers, and Ornish diet) found that weight change did not significantly differ between

participants assigned to any of the four diets, and better dietary adherence predicted higher amount of weight loss. The most difficult diets for participants to adhere to were the very low-fat diet (Ornish) and the low-carbohydrate diet (Atkins). In a recent meta-analysis of five dietary trials for weight loss, low-carbohydrate non–energy-restricted diets were found to be as effective as low-fat, energy-restricted diets for weight loss up to 1 year. In a meta-analysis of 48 randomized controlled trials testing a variety of named diets to no diet for weight loss, there was no difference in weight loss from the low carbohydrate diet trials compared to the low fat diet studies (both types of diet had mean −7.3 kg of weight loss at 12 months vs no diet).

Physical activity interventions   The Institute of Medicine recommends physical activity of 60 min/d for most days of the week for weight loss and/or control of weight. The recommendations for heart disease prevention are 30 min/d of moderate-intensity physical activity on most days per week, or at least 150 min/wk.

In a systematic review of 21 weight loss trials of diet with or without an exercise component, there was moderate quality evidence that both diet and exercise together was most effective for obesity management. Head-to-head trials testing the efficacy of weight loss with diet alone versus exercise alone have found that diet is associated with greater weight loss (−2.9 kg on average) than exercise alone. Isolated aerobic exercise for weight loss was associated with an average of 1.6 kg of weight loss in another meta-analysis.

Exercise has been shown to be important for the maintenance of weight loss. In a randomized controlled trial of 140 postmenopausal women who had lost 25 lbs with a very low calorie diet, all measures of energy expenditure declined after weight loss, but the groups randomized to either aerobic or resistance training had increases in total energy expenditure and non-exercise associated thermogenesis compared to a group that did not exercise.

Pharmacologic intervention/CAM   Weight loss medications have had mixed results with most delivering modest amounts of short-term weight loss, often associated with many undesirable side effects. Currently there are six main categories of weight loss medications approved by the FDA: adrenergic agents, serotonergic agents, combination of both adrenergic/serotonergic agents, lipase inhibitors, GABA-receptor antagonists, and GLP-1 receptor agonists. A majority of available obesity treatments work by suppressing appetite centrally. One approved medication, orlistat, prevents digestion and absorption of dietary fat by inhibiting the gut enzyme lipase.

Table 20–5 shows specific medications in each category, the amount of weight loss reported in trials compared to placebo from a recent review and meta-analysis, and the common side effects of each agent. Lorcaserin, an agonist that targets serotonin 2C receptors within the hypothalamus, promotes the release of melanotropin-α, which stimulates melanocortin receptor-4, resulting in appetite satiety. A phentermine/topiramate combination medication targets multiple pathways that regulate energy homeostasis and appears to be the most effective obesity medication available, although long-term studies are needed. The mechanism of action of naltrexone/bupropion is not completely understood but includes the hypothalamic melanocortin system and mesolimbic reward system. Liraglutide, a glucagon-like peptide (GLP)-1 agent, works peripherally by decreasing gastric emptying and in the caudal brainstem to improve satiety.

Complementary and alternative medications such as herbs, vitamins, nutritional supplements, and meal replacement therapies are commonly used by the general public for weight loss. However, the efficacy and safety of these herbal remedies have not been well studied. A recent meta-analysis of trials found that compounds containing ephedra, *Cissus quadrangularis*, ginseng, bitter melon, and zingiber had some short-term efficacy with weight loss with mostly mild adverse effects. However, longer-term randomized controlled trials are needed to prove efficacy and safety.

TABLE 20–5 Medications for treatment of obesity.

Drug	Category	Recommended Dose	Mean Weight Loss vs Placebo	Side Effects
Phentermine	Adrenergic agent	15-37.5 mg daily	3.6 kg (0.6-6.0 kg) greater weight loss than placebo	Dry mouth, headache, insomnia, irritability, tachycardia, ↑BP
Orlistat	Lipase inhibitor	60-120 mg 3 times daily	-5.8 kg vs -3.0 kg for placebo	Diarrhea, flatulence, fecal urgency and incontinence, abdominal pain, dyspepsia
Lorcaserin	Selective serotonin 2c receptor agonist	10 mg daily or twice daily	-4.5% to -5.8% vs -1.5% to -2.5% for placebo	Headache, nausea, dizziness, fatigue, dry mouth, hypoglycemia
Phentermine/topiramate	Adrenergic agent/GABA receptor antagonist	7.5/46 mg daily	-10.9% vs -1.6% for placebo	Paresthesias, dry mouth, constipation, insomnia, nausea, depression
Naltrexone/bupropion SR	Opioid receptor antagonist/Adrenergic agent	90 mg/8 mg to max dose of 360 mg/32 mg	-5.0% to -9.3% vs -1.2% to -5.1% in placebo	Nausea, headache, constipation, dizziness, insomnia, dry mouth
Liraglutide	GLP-1 receptor agonist	0.6 mg SC/day, to max dose 3 mg/day	-6.2% to -8.0% vs -0.2% to -2.6% in placebo	Nausea, vomiting, diarrhea, constipation, hypoglycemia, upper respiratory tract infection

B. Bariatric surgery   Surgical treatment for weight loss has been performed for the past 50 years, and these procedures are increasing in popularity. Approximately 20,000 weight-loss surgeries were performed in 1995 and over 179,000 were done in 2013, a ninefold increase. There are three main types of bariatric surgeries: restrictive, malabsorptive, and those that have both a restrictive and a malabsorptive component. Purely restrictive surgeries include horizontal gastroplasty, vertical banded gastroplasty, silastic ring vertical gastroplasty, and adjustable gastric banding. Purely malabsorptive surgeries include the oldest of all bariatric procedures, the jejunoileal bypass, biliopancreatic diversion, duodenal switch, and long limb gastric bypass. The combination of restrictive and malabsorptive surgical procedure is the most commonly used bariatric surgery worldwide, the Roux-en-Y gastric bypass procedure (RYGB), which involved stapling the upper stomach into a small 30-mL pouch and attaching it to the jejunum bypassing the lower stomach and duodenum. This procedure can be done with an open surgical approach, or now more commonly with a laparoscopic approach.

Several studies have examined the effects of different types of bariatric surgeries on various outcomes including weight loss, comorbid disease, complication rates, and mortality. There are very few randomized controlled trials comparing different surgical procedures, so most of the outcomes data come from case series, prospective cohort analyses, or nonrandomized trials. The largest and longest double cohort study to date is the Swedish Obese Subjects (SOS) study in which obese adults who voluntarily underwent bariatric surgery were matched to a control group of medically treated patients. After 10 years of follow-up of 1703 participants, those treated with surgery had a 16% weight loss compared to 2% in the controls ($p < 0.001$). They compared the different types of bariatric surgery approaches in this study and found that those who had gastric bypass lost more weight than those who received banding procedures or the vertical banded gastroplasty.

An earlier meta-analysis determined the effect of bariatric surgery on weight loss, mortality, and four comorbid diseases including diabetes, hyperlipidemia, hypertension, and sleep apnea. The mean excess weight loss was 61.2% (58.1%-64.4%), highest in those who had the biliopancreatic diversion or duodenal switch (70.1%) and lowest in those who underwent gastric banding (47.5%). Operative mortality within 30 days of the surgery was 0.1% for purely restrictive procedures and 0.5% for gastric bypass, and 1.1% for biliopancreatic diversion or duodenal switch procedures. Type 2 diabetes resolved in approximately 77% of patients after surgery, hyperlipidemia improved in 70%, hypertension resolved in 62%, and OSA resolved in 86%.

Much attention has focused on determining the mechanisms involved in the dramatic weight loss associated with bariatric surgery. Previous hypotheses that restrictive procedures reduce the quantity of food ingested at any one time, malabsorptive procedures cause wasting of fat calories, with RYGB employing a combination of both mechanisms, have not been supported by studies. More recent thinking attributes weight loss after bariatric procedures to a substantial decline in hunger and increase in satiety that is regulated through neuroendocrine mechanisms and gut hormones such as ghrelin and PYY. The effect of RYGB surgery on ghrelin levels is controversial since some studies have found increased levels, no change, or decreased levels after surgery. These inconsistent results may be explained by the timing of ghrelin sampling after surgery. After purely restrictive surgeries, ghrelin levels increase, which may help explain the more dramatic weight loss observed following RYGB versus purely restrictive surgeries. PYY levels rise to those of a nonobese person following vertical banded gastroplasty. However, with RYGB, there is an early exaggerated response in PYY secretion, approximately two- to fourfold greater than that observed in lean, obese, or gastric banded patients, which may contribute to the sustained weight loss seen with this type of procedure.

Defining appropriate patient criteria to minimize risks and maximize the benefits from bariatric surgery have been debated. Some studies have concluded that surgical intervention is more effective for weight loss and control of comorbid diseases than nonsurgical treatments in patients with a BMI greater than or equal to 40 kg/m^2. In 2004, with extensive scientific input, the National Coverage Advisory Committee panel concluded that bariatric surgery could be offered to Medicare beneficiaries with BMI greater than or equal to 35 kg/m^2 who have at least one comorbid disease associated with obesity and have been unsuccessful previously with medical treatment of obesity. Other criteria for patient selection proposed by leading associations include adequate patient commitment with medical follow-up and use of dietary supplements, no other reversible endocrine disorders causing obesity, no current substance abuse, and no severe psychiatric illness.

C. Interventions with efficacy in children   Several studies have investigated behavioral and lifestyle intervention programs for pediatric weight loss. Most randomized controlled trials of overweight children and adolescents have found positive, but small-to-moderate effects, of combined lifestyle interventions on BMI. Family-based behavioral weight loss programs have produced larger effects that persist for several months of follow-up. Less is known about the long-term safety and efficacy of pharmacologic agents for treatment of pediatric obesity.

REFERENCES

Barlow SE, Dietz WH. Obesity evaluation and treatment: expert committee recommendations. The Maternal and Child Health Bureau, Health Resources and Services Administration and the Department of Health and Human Services. *Pediatrics*. 1998;3:E29. [PMID: 1224658]

Berthoud H-R, Morrison C. The brain, appetite, and obesity. *Ann Rev Psychol*. 2008;59:55. [PMID: 18154499]

Buchwald H, Avidor Y, Braunwald E, et al. Bariatric surgery: a systematic review and meta-analysis. *JAMA*. 2004;14:1724. [PMID: 15479938]

CDC Overweight & Obesity. https://www.cdc.gov/obesity/index.html. Accessed June 14, 2016.

Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med*. 2007;4:370. [PMID: 17652652]

Cummings DE, Overduin J. Gastrointestinal regulation of food intake. *J Clin Invest*. 2007;1:13. [PMID: 17200702]

Dansinger ML, Gleason JA, Griffith JL, et al. Comparison of the Atkins, Ornish, Weight Watchers, and Zone diets for weight loss and heart disease risk reduction: a randomized trial. *JAMA*. 2005;1:43. [PMID: 15632335]

Dickerson LM, Carek PJ. Pharmacotherapy for the obese patient. *Prim Care*. 2009;2:407. [PMID: 19501251]

Galic S, Oakhill JS, Steinberg GR. Adipose tissue as an endocrine organ. *Mol Cell Endocrinol*. 2010;2:129. [PMID: 19723556]

Grundy SM, Cleeman JI, Daniels SR, et al. Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation*. 2004;3:433. [PMID: 14744958]

Hedley AA, Ogden CL, Johnson CL, et al. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002. *JAMA*. 2004; 23:2847. [PMID: 15199035]

Hsu WC, Araneta MRG, Kanaya AM, et al. BMI cut points to identify at-risk Asian Americans for type 2 diabetes screening. *Diabetes Care*. 2015;38:150. [PMCID: PMC4392932]

Hunter GR, Fisher G, Neumeier WH, et al. Exercise training and energy expenditure following weight loss. *Med Sci Sports Exerc*. 2015;47:1950.

Johnston BC, Kanters S, Bandayrel K, et al. Comparison of weight loss among named diet programs in overweight and obese adults: a meta-analysis. *JAMA*. 2014;312:923.

Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*. 2006;7121:840. [PMID: 17167471]

Kenchaiah S, Evans JC, Levy D, et al. Obesity and the risk of heart failure. *N Engl J Med*. 2002;5:305. [PMID: 12151467]

Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. *Am J Clin Nutr*. 2007;3:556. [PMID: 17823417]

Montague CT, O'Rahilly S. The perils of portliness: causes and consequences of visceral adiposity. *Diabetes*. 2000;6:883. [PMID: 10866038]

Morton GJ, Cummings DE, Baskin DG, et al. Central nervous system control of food intake and body weight. *Nature*. 2006;7109:289. [PMID: 16988703]

Nordmann AJ, Nordmann A, Briel M, et al. Effects of low-carbohydrate vs low-fat diets on weight loss and cardiovascular risk factors: a meta-analysis of randomized controlled trials. *Arch Intern Med*. 2006;3:285. [PMID: 16476868]

Ogden CL, Carroll MD, Fryar CD, Flegal KM. Prevalence of obesity among adults and youth: United States, 2011-2014. *NCHS Data Brief*. 2015;219:1.

Ornellas T, Chavez B. Naltrexone SR/Bupropion SR (Contrave): a new approach to weight loss in obese adults. *Pharmacy and Therapeutics*. 2011;36:255.

Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes*. 1988;12:1595. [PMID: 3056758]

Romero-Corral A, Montori VM, Somers VK, et al. Association of bodyweight with total mortality and with cardiovascular events in coronary artery disease: a systematic review of cohort studies. *Lancet*. 2006;9536:666. [PMID: 16920472]

Schwingshackl L, Dias S, Hoffmann G. Impact of long-term lifestyle programmes on weight loss and cardiovascular risk factors in overweight/obese participants: a systematic review and network meta-analysis. *Syst Rev*. 2014;3:130.

Sjostrom L, Narbro K, Sjostrem CD, et al. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med*. 2007;8:741. [PMID: 17715408]

Sweeting AN, Hocking SL, Markovic TP. Pharmacotherapy for the treatment of obesity. *Mol Cell Endocrinol*. 2015;418 Pt 2:173.

Thorogood A, Mottillo S, Shimony A, et al. Isolated aerobic exercise and weight loss: a systematic review and meta-analysis of randomized controlled trials. *Am J Med*. 2011;124:747.

Van Gaal LF, Mertens IL, De Block CE. Mechanisms linking obesity with cardiovascular disease. *Nature*. 2006;7121:875. [PMID: 17167476]

WHO. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organization Technical Report Series. 2000;1.

This page intentionally left blank

Humoral Manifestations of Malignancy

Dolores M. Shoback, MD and Janet L. Funk, MD

ACTH	Adrenocorticotropic hormone		**IL**	Interleukin
ANP	Atrial natriuretic peptide		**LH**	Luteinizing hormone
APUD	Amine precursor uptake and decarboxylation (cell)		**MEPE**	Matrix extracellular phosphoglycoprotein
CRH	Corticotropin-releasing hormone		**MRI**	Magnetic resonance imaging
CT	Computed tomography		**NET**	Neuroendocrine tumor
FDG-PET	Fluorodeoxyglucose positron emission tomography		**PACAP**	Pituitary adenylate cyclase activating peptide
FGF	Fibroblast growth factor		**PHEX**	Phosphate-regulating gene with homologies to endopeptidases on the X chromosome
FSH	Follicle-stimulating hormone			
GH	Growth hormone		**POMC**	Proopiomelanocortin
GHRH	Growth hormone–releasing hormone		**PTH**	Parathyroid hormone
hCG	Human chorionic gonadotropin		**PTHrP**	Parathyroid hormone-related protein
HTLV-1	Human T cell leukemia virus-1		**SIADH**	Syndrome of inappropriate antidiuretic hormone (secretion)
IGF	Insulin-like growth factor		**TNF**	Tumor necrosis factor
IGF-BP3	Insulin-like growth factor binding protein-3		**VIP**	Vasoactive intestinal peptide

ECTOPIC HORMONE AND RECEPTOR SYNDROMES

Some of the most challenging endocrine problems occur in patients with malignancies of diverse cell types, because both endocrine and nonendocrine tumors secrete polypeptide hormones. As it became recognized that a polypeptide hormone could be produced by tumor cells derived from a tissue that normally did not secrete the hormone, the notion of ectopic hormone production developed. Most tumors associated with ectopic hormone syndromes are derived from cells that are normally capable of producing peptide hormones. Initially, it was thought that ectopic hormone production by tumor cells was a rare event. Interestingly, both the frequency and the original conception of this syndrome have been redefined over the last few decades. It has come to be appreciated—through the use of biochemical and molecular biologic techniques—that the synthesis of peptide hormones and the transcription of their genes by tumor cells are in fact quite common occurrences. Tumor cells may differ from normal cells in their ability or inability to process precursor molecules, which may account for the presence or absence of hormone excess states and for the profile of peptide hormone forms and fragments present in the circulation and in tumor cell extracts. However, tumor production of hormone fragments or precursors is much more common than the clinical syndromes of hormone excess.

The classic criteria used to confirm that a tumor is the source of a hormone excess state include the following: (1) evidence of an endocrinopathy in a patient with a tumor; (2) remission of the endocrinopathy after tumor resection; (3) detection of an arteriovenous gradient across the tumor; and (4) confirmation of protein and messenger RNA encoding the hormone produced by tumor tissue.

In addition to classic hormone excess states resulting from the ectopic or inappropriate secretion of a hormone by an endocrine or nonendocrine tumor, endocrinopathies can result from the ectopic expression of a hormone's receptor. This is well illustrated, for example, by the occurrence of Cushing syndrome in pregnancy or in relation to meals, due to the ectopic expression of luteinizing hormone (LH) or gastric inhibitory polypeptide (GIP) receptors in adrenal tissue, respectively. Several other examples of ectopic receptor syndromes have been documented. Some of these will be discussed later, particularly as a cause for unusual forms of adreno-corticotropic hormone (ACTH)-independent Cushing syndrome.

A variety of peptides are produced by both benign and malignant tumors as listed in Table 21–1. The biochemical pathways and machinery leading to the synthesis, processing, and secretion into the circulation of a peptide hormone are present in all cells. In contrast, the multiple enzymatic steps that lead to the production of a highly active steroid hormone (eg, cortisol or 1,25-dihydroxyvitamin D [1,25-(OH)$_2$ D]) are restricted, with rare exceptions, to steroid-producing cells or their precursor cells. Hence, the occurrence of 1,25-(OH)$_2$ D excess caused by a tumor is distinctly unusual, being observed infrequently in hematologic malignancies and only in those with the ability to 1-hydroxylate 25-OH D, the immediate precursor of 1,25-(OH)$_2$ D. Ectopic hormone syndromes—the most common of the paraneoplastic syndromes—thus predominantly reflect peptide hormone excess states. The most common ectopic peptide hormone syndromes are described in greater detail in subsequent sections of this chapter.

TABLE 21–1 Polypeptide hormones produced ectopically by benign and malignant tumors and their associated endocrinopathies.

Hormone	Syndrome
Parathyroid hormone-related protein (PTHrP)	Hypercalcemia
Parathyroid hormone (rare)	Hypercalcemia
Adrenocorticotropin (ACTH)	Cushing syndrome
Corticotropin-releasing hormone (CRH)	Cushing syndrome
Antidiuretic hormone (ADH)	Hyponatremia
Insulin and insulin-like growth factors (IGF)	Hypoglycemia
Growth hormone–releasing hormone (GHRH)	Acromegaly
Calcitonin	*No specific syndrome*
Human chorionic gonadotropin (hCG)	Children: precocious puberty Men: erectile dynsfunction, gynecomastia Women: dysfunctional uterine bleeding
Fibroblast growth factor (FGF) 23	Oncogenic osteomalacia
Vasoactive intestinal peptide (VIP)	Diarrhea

APUD Concept of Neuroendocrine Cell Tumors

Over the years since the initial recognition that nonendocrine tumors were the source of the ectopic hormones produced in these endocrine syndromes, the notion developed that the hormones originated from highly specialized neuroendocrine cells in tumors. These cells were thought to derive from the neural crest and were postulated to be able to synthesize and store biogenic amines and were thus designated as amine precursor uptake and decarboxylation (APUD) cells. Neuroendocrine cells, like calcitonin-secreting C cells and adrenal chromaffin cells, clearly had these properties, and tissues giving rise to ectopic hormone syndromes (eg, lung and gastrointestinal tract) also had APUD cells scattered throughout them. It was originally thought that the tumor cells producing excessive amounts of polypeptide hormones were derived exclusively from APUD cells in the tissue of origin (eg, ACTH-producing cells of the lung).

Newer insights into tumor cell biology have led to a better understanding of the pathogenesis and etiology of the ectopic hormone production states. Studies have shown that not all APUD cells are derived from the neural crest. Some of these cells are of endodermal origin. Furthermore, with careful examination of tumors, it has become clear that peptide hormones are often produced by non-APUD cells. It has, however, been appreciated that the clinically evident hormone hypersecretion states are typically caused by tumors that are in fact derived from APUD cells. These cells are the ones with full capacity to produce and store peptide hormones efficiently in dense secretory granules and then release biologically significant quantities of active hormones into the circulation.

HYPERCALCEMIA OF MALIGNANCY

Hypercalcemia is the most common paraneoplastic endocrine syndrome, occurring in approximately 20% of patients with cancer. In the great majority of patients (98%), the identity of the tumor is apparent at the time of presentation, and the prognosis is poor, as most patients with hypercalcemia of malignancy do not survive beyond 6 months. As discussed more fully in Chapter 8, signs and symptoms of hypercalcemia include nausea, vomiting, severe dehydration (exacerbated by hypercalcemia-induced nephrogenic diabetes insipidus) and a change in mental status.

Pathogenesis

Elevated serum calcium levels primarily result from excessive osteoclast-mediated bone resorption. Tumor-driven osteolysis occurs via two distinct mechanisms: (1) humoral effects of systemically elevated tumor-derived factors and/or (2) localized osteolysis adjacent to malignant cells within the bone marrow (Figure 21–1). When the syndrome of hypercalcemia of malignancy was first appreciated, direct invasion of bone by osteolytic tumors was thought to be causative. However, in the 1980s, a humoral basis was identified as the most frequent cause (>80%), even in settings where focal lytic bone metastases were present.
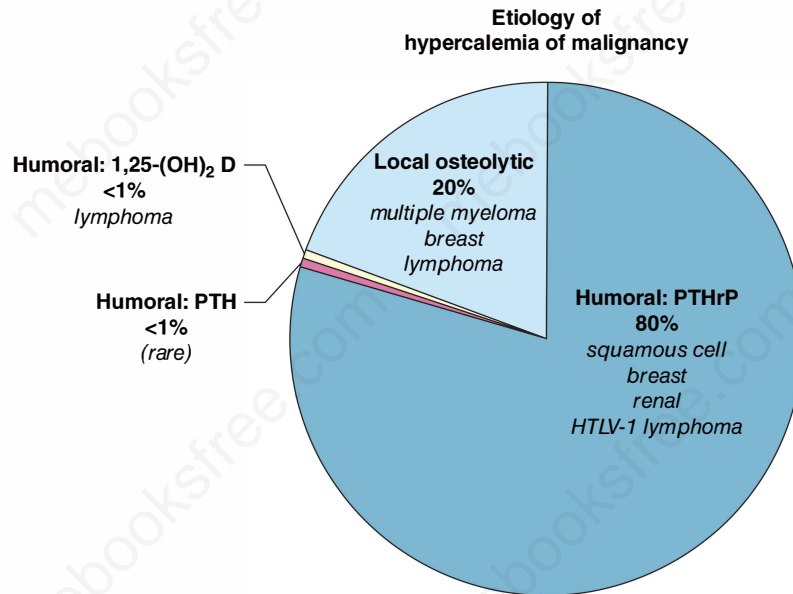
FIGURE 21–1 Schematic representation of causes and tumor types associated with hypercalcemia of malignancy.

Decreased renal calcium excretion may also contribute to pathogenesis, either because of the hypocalciuric effects of certain humoral mediators of hypercalcemia, such as parathyroid hormone (PTH)-related protein (PTHrP) (discussed later), or because of the decreased glomerular filtration that occurs with hypercalcemia-induced nephrogenic diabetes insipidus.

Humoral Mediators

A humoral cause of hypercalcemia of malignancy was first proposed by Fuller Albright in the 1940s when describing a case occurring in the absence of significant bone metastases. It would take another 50 years for this humoral factor to be identified. It was appreciated early on that the basic biochemical parameters of this syndrome (high serum calcium, low serum phosphorus, elevated nephrogenous cyclic adenosine monophosphate or cAMP) were suggestive of ectopic PTH production. However, PTH was not the culprit, as serum PTH levels in these cases were actually (and appropriately) suppressed. Ultimately, and unlike most other paraneoplastic endocrine syndromes that are caused by well-described hormones of known function, a novel peptide, PTH-related protein (PTHrP), was identified as the humoral mediator of this syndrome in a majority of cases (see Figure 21–1).

PTHrP, initially so named because of its shared biochemical/metabolic effects with PTH, has much in common with PTH. The amino terminal portion of PTHrP bears strong homology to PTH, such that the two peptides bind with similar affinity to a common receptor (now known as the PTH/PTHrP type 1 receptor [PTHR1]) in bone and kidney. Therefore, the biochemical markers of PTHrP-mediated hypercalcemia *in vivo*, when unregulated production by tumors causes elevated circulating PTHrP levels, are similar to those of hyperparathyroidism. Some unexplained differences, however, are seen in PTHrP-mediated

hypercalcemia of malignancy (vs primary hyperparathyroidism), including normal or suppressed 1,25-(OH)$_2$ D levels and an uncoupling of bone resorption and formation that results in severe bone loss. The reasons for these differences are not understood but are speculated to include: (1) the ability of profound hypercalcemia per se to suppress 1,25-(OH)$_2$ D levels; (2) the contributions of additional tumor-derived cytokines, such as interleukin (IL)-1α or IL-6, to the process of bone resorption; and (3) the ability of PTH and PTHrP to interact with the PTHR1 and activate signal transduction pathways differently.

Subsequent study of this previously unknown peptide demonstrated that PTHrP and PTH arose from ancient duplication of a common gene and diverged in their normal functions in many critical ways. Each hormone evolved separately such that in higher vertebrates, PTH regulates calcium homeostasis in an endocrine manner, while PTHrP has critical developmental roles, acting locally in a paracrine or autocrine fashion via binding to PTHR1-expressing cells, to mediate endochondral bone growth, tooth eruption, and the development of the mammary gland and cardiovascular system. As with other developmentally regulated proteins, PTHrP plays less of a role in adult homeostasis, but can be reexpressed in response to specific changes in gene programming, including pregnancy (calcium regulation by the lactating mammary gland and placenta), injury and inflammation (regulation of vascular tone in ischemia, sepsis, and inflammation-associated bone resorption), and/or tumorigenesis (hypercalcemia of malignancy).

When tumor expression of PTHrP results in inappropriately high PTHrP levels that reach bone cells via the circulation or local synthesis in the bone microenvironment, a vicious cycle can ensue. PTHrP stimulates osteoblast or stromal cell expression of RANKL (receptor activator of nuclear factor kappa B ligand), the primary gatekeeper modulating bone resorption in health and disease. RANKL binds to its receptor, RANK, on osteoclasts and

their precursors, stimulating osteoclast differentiation and function. Increased numbers of bone-resorbing osteoclasts are generated both by the local release of PTHrP in the case of bone metastases or by high systemic levels of tumor-derived PTHrP. Both mechanisms cause enhanced bone resorption. In the case of bone metastases, sequestered growth factors, such as transforming growth factor (TGF)-beta, released locally from the bone matrix during resorption, are thought to further enhance tumor cell secretion of PTHrP.

Although PTHrP is by far the most common humoral mediator of hypercalcemia of malignancy (see Figure 21–1), other calciotropic hormones can also cause the syndrome (<1% of cases). Elevated circulating 1,25-(OH)$_2$ D levels, which enhance intestinal calcium absorption and stimulate osteoclast differentiation, can cause hypercalcemia in lymphoma due to increased 1α-hydroxylase activity in tumor-associated macrophages and possibly the lymphoproliferative cells themselves. Ectopic production of PTH is extremely rare, but has occasionally been reported for neuroendocrine tumors (NETs) and other solid tumors.

Solid Tumors Associated with Hypercalcemia of Malignancy

The majority of cases of hypercalcemia occurring with solid tumors (80%) are associated with two common tumor types, squamous cell lung carcinoma and breast carcinoma, and are primarily due to humoral effects of PTHrP (see Figure 21–1). PTHrP-induced hypercalcemia is also frequently seen in renal cell carcinoma. Indeed, PTHrP was first isolated and identified by independent research groups using each of these three tumor types. In contrast, hypercalcemia is rarely seen in other commonly occurring cancers (eg, colon, gastric, thyroid, and small cell lung carcinomas), including tumors that are frequently metastatic to bone (eg, prostate cancer).

A. Squamous cell carcinomas   Squamous cell carcinomas account for more than one-third of all cases of hypercalcemia of malignancy. Humoral effects of tumor-derived PTHrP account for most cases of hypercalcemia in this setting, even when bone metastases are present. Twenty-five percent of patients with squamous cell lung carcinomas develop PTHrP-mediated hypercalcemia, and other sites of squamous cell carcinoma (head, neck, esophagus, cervix, vulva, and skin) are similarly associated with a high incidence of hypercalcemia.

B. Breast carcinoma   While osteolytic bone metastases are a hallmark of advanced breast cancer, occurring in 70% of cases, the majority of these women do not have hypercalcemia. However, when hypercalcemia does occur (30% of advanced-stage breast cancer), it is rarely seen in the absence of bone metastases. Most breast cancer bone metastases (>92%) express PTHrP; thus a local increase in PTHrP is thought to be a key driver of osteolysis-induced hypercalcemia (see Figure 21–1). In addition, humoral effects of PTHrP, as evidenced by elevated circulating levels of PTHrP associated with increased nephrogenous cAMP, have been

documented in two-thirds of hypercalcemia cases occurring with bone metastases as well as in the majority of cases occurring in the absence of bone metastases. Hypercalcemia seen in late stages of breast cancer is usually unremitting and associated with a survival time of weeks to months. Despite this close association of PTHrP with advanced disease, the prognostic significance and pathophysiologic role of PTHrP in primary breast carcinomas, two-thirds of which have been reported to express PTHrP, remain unclear.

C. Renal cell carcinoma   Renal cell carcinoma, while less common than squamous cell or breast carcinoma, is frequently associated with hypercalcemia in advanced cases. While bone metastases are common in advanced disease, hypercalcemia appears to be due primarily to humoral effects of PTHrP. In some series, 100% of renal cell tumors have been reported to stain positively for immunoreactive PTHrP. However, hypercalcemia is reported in less than 20% of cases.

D. Other solid tumors   Hypercalcemia is also associated with other solid tumors, although less frequently. In most of these cases, humoral effects of PTHrP are causative. Bladder carcinoma, large cell and adenocarcinoma of the lung, and endocrine tumors (including islet cell tumors, pheochromocytoma, and carcinoid tumors) have also been reported to secrete PTHrP. PTHrP-mediated hypercalcemia has also been reported for ovarian carcinomas. There are also rare case reports of ectopic PTH production by various neoplasms (eg, ovarian carcinoma, gastric carcinoma, lung cancer [small and squamous]).

Hematologic Malignancies Associated with Hypercalcemia of Malignancy

A. Multiple myeloma   Hypercalcemia (present in 10%-30% at diagnosis) and lytic bone lesions (present in up to 70% at diagnosis) are among the diagnostic criteria for multiple myeloma (MM), which is one of the most common hematologic malignancies. Focal oseolysis, the cause of hypercalcemia in MM (see Figure 21–1), occurs adjacent to malignant plasma cells, which infiltrate the bone marrow and secrete multiple factors capable of stimulating osteoclast differentiation and activity. A major effect of these factors, of which macrophage inflammatory peptide-1 is a key example, is to induce RANKL expression in cells of the osteoblastic lineage. Evidence of RANKL expression by plasma cells also exists. Locally increased RANKL causes bone lysis by binding to its cognate receptor RANK on osteoclast lineage cells, stimulating their differentiation and activity. Defects in bone formation, due to suppression of the Wnt signaling pathway by other tumor factors (eg, Dickkopf-1), also accompany lytic bone disease in MM. In 30% of cases, bone lysis leads to hypercalcemia. Because renal disease occurs frequently in MM due to the filtration of Bence Jones proteins (light chain fragments of IgG), it is hypothesized that patients with renal impairment may be particularly predisposed to development of hypercalcemia in this setting of increased bone resorption.

B. Lymphoma and leukemia Hypercalcemia has been reported in up to 15% of patients with lymphoma; is seen primarily in patients with bone involvement; and can occur with a variety of cell types. In most cases, the etiology is evenly divided between: (1) local lytic effects of tumor-derived factors within the marrow, analogous to the mechanisms discussed for MM, or (2) humoral effects of tumor-derived 1,25-$(OH)_2$ D (see Figure 21–1). 1,25-$(OH)_2$ D-mediated hypercalcemia of malignancy is nearly unique to lymphoma. Its underlying pathology is analogous to that seen in hypercalcemia in granulomatous disorders; in both settings, 1α-hydroxylase activity in infiltrating macrophages is thought to result in unregulated conversion of 25-(OH) D to active 1,25-$(OH)_2$ D. Increased intestinal calcium absorption and increased bone resorption contribute to hypercalcemia in this setting. In addition, there are numerous case reports of humoral effects of tumor-derived PTHrP as a cause of hypercalcemia in non-Hodgkin lymphomas (particularly B cell), acute lymphoblastic leukemia, and blast crises in chronic myeloid leukemia. Patients diagnosed with adult T cell leukemia/lymphoma induced by HTLV-1 (human T cell lymphotrophic virus-1), the first identified human retrovirus, must be considered separately when evaluating possible causes of hypercalcemia. In these patients, hypercalcemia occurs in two-thirds of cases, responds poorly to treatment, and is due to the humoral effects of tumor-derived PTHrP. Tax1, a transcription factor encoded by the HTLV-1 genome, transforms T cells in these cases by altering the regulation of hundreds of host genes, including transactivation of PTHrP. Thus, the clinician confronting lymphoma-induced hypercalcemia must be cognizant of the multiple etiologies driving hypercalcemia in that setting.

Diagnosis

The diagnosis of hypercalcemia is discussed in detail in Chapter 8. Primary hyperparathyroidism, the most common cause of hypercalcemia in outpatients, must still always be considered as a potential cause of hypercalcemia in patients with malignancy, where it can be causal in approximately 10% of patients and is simply screened for by assaying PTH levels. In the setting of hypercalcemia of malignancy with normal renal function, PTH levels are appropriately suppressed. Further evaluation can be guided, in part, by the tumor type, which is usually known prior to presentation with hypercalcemia. Elevated PTHrP is detected in a majority of cases of hypercalcemia of malignancy associated with solid tumors, including breast carcinoma with or without bone metastases, and in HTLV-1–induced T cell lymphomas. Measurement of 1,25-$(OH)_2$ D should be considered in all cases of lymphoma-associated hypercalcemia. Lytic bone lesions are usually identified during tumor staging.

Treatment

The treatment of hypercalcemia is discussed in detail in Chapter 8. In addition to an important role for saline rehydration to promote calciuresis, because bone resorption is central to all causes of hypercalcemia of malignancy, osteoclast-targeted bisphosphonates are a mainstay of therapy. Calcitonin is sometimes also used in the acute setting as its effect is more rapid, albeit of limited magnitude and duration. Denosumab, a monoclonal antibody targeting RANKL that prevents osteoclast differentiation, survival, and resorptive function, has recently been approved for use in the United States in cases of bisphosphonate-refractory hypercalcemia. Glucocorticoids may be efficacious in the treatment of hypercalcemia due either to the local lytic effects of neoplastic plasma cells in MM or to the increased production of 1,25-$(OH)_2$ D in lymphoma. Both bisphosphonates and denosumab are becoming mainstays of palliative care for patients with hypercalcemia; both have demonstrated ability to decrease skeletal-related events (metastases and pathologic fractures) and ameliorate bone pain in these patients. Possible use of these agents for the primary prevention of bone metastases (not yet established) in breast and other cancers is currently an active area of clinical research.

ECTOPIC CUSHING SYNDROME

Many malignant tumors produce the ACTH precursor proopiomelanocortin (POMC) but lack the enzymes necessary to process this precursor to biologically active ACTH. Therefore, only a fraction of such tumors release sufficient ACTH to cause Cushing syndrome. Initially, the tumors recognized to cause this syndrome were of nonpituitary but endocrine origin, such as islet cell carcinomas and pheochromocytomas. Subsequently, a wide variety of different tumor cell types, both endocrine and nonendocrine, have been associated with the ectopic ACTH syndrome.

The classic description of the ectopic ACTH syndrome was made by Grant Liddle and coworkers in the early 1960s and was based on a series of patients who mostly had highly malignant tumors (eg, small cell carcinoma of the lung). More recently, the ectopic ACTH syndrome has been recognized with increasing frequency with benign tumors, specifically carcinoids. Benign lesions typically present in a more subtle clinical manner, often over months to years, before the tumor is identified. The more gradual development of the clinical syndrome plus the more subtle biochemistry have led to considerable challenges in distinguishing this form of the ectopic ACTH syndrome from pituitary tumors causing Cushing disease. This subtle variant of tumor-induced ACTH excess has been dubbed the *occult* ectopic ACTH syndrome, especially when the tumor is not evident by the best diagnostic imaging modalities. Tumors can also cause an ectopic ACTH-like syndrome through production of corticotropin-releasing hormone (CRH), albeit rarely. Indeed, some of the tumors that make the latter cosecrete ACTH as well. Ectopic CRH production has been seen in bronchial carcinoids, medullary thyroid carcinoma, and metastatic prostatic cancer.

Differential Diagnosis

Cushing syndrome—signs and symptoms resulting from unregulated production of glucocorticoids—is caused by a number of underlying disturbances. These must be differentiated to ensure successful treatment. Causes include pituitary ACTH-dependent

Cushing disease, adrenal tumors or ACTH-independent Cushing syndrome, and the ectopic ACTH syndrome. In several large series, 50% to 80% of patients with Cushing syndrome have a pituitary cause. Adrenal adenomas (and very rarely carcinomas) account for 5% to 30% of cases of Cushing syndrome. The ectopic ACTH syndrome comprises approximately 10% to 20% of cases of Cushing syndrome.

A wide variety of tumors cause ectopic ACTH syndrome (Table 21–2). In the classic and initial descriptions of this syndrome, there was a preponderance of malignant tumors, particularly small cell carcinomas of the lung. It is now clear that most cases of ectopic ACTH syndrome are due to benign tumors. Most recently, microscopic carcinoid *tumorlets*, particularly in the lung, have been recognized to cause occult ectopic ACTH syndrome. These tumors may be exceptionally difficult to diagnose by standard techniques.

The diagnosis of Cushing syndrome requires a rigorous approach. Cushing syndrome should be suspected first on solid clinical grounds and then established biochemically. This is accomplished by demonstrating the presence of hypercortisolism—a frankly elevated 24-hour urinary free cortisol level, lack of suppression of plasma cortisol levels after a 1-mg overnight dexamethasone suppression test, or elevated late night salivary cortisol levels (see Chapter 9). In Cushing syndrome due to any cause and often in ectopic ACTH syndrome, random cortisol levels are elevated. Once hypercortisolism is established, plasma ACTH levels are measured. These levels are markedly elevated in classic forms of the ectopic ACTH syndrome, typically secondary to malignant lung neoplasms. There is, however, considerable overlap between the milder cases of the ectopic ACTH syndrome, caused by benign and slowly growing tumors, and Cushing disease due to a pituitary tumor. In the former case, the tumors are often small and clinically silent—hence the descriptor *occult* ectopic ACTH syndrome. For these reasons, rigorous biochemical criteria must be applied in appropriate clinical situations to make certain that the correct diagnosis is made. Plasma ACTH levels in patients with clinically evident tumors are often strikingly elevated (390-2300 pg/mL [87-511 pmol/L] by radioimmunoassay). Individuals with ectopic ACTH syndrome due to occult tumors have ACTH levels that overlap with pituitary-dependent Cushing disease (42-428 pg/mL [9.3-95 pmol/L]). It is said that patients with plasma ACTH levels greater than 200 pg/mL (44.4 pmol/L) typically have the ectopic ACTH syndrome, although further testing is required to prove this and to localize the tumor.

After hypercortisolism and ACTH excess are established, the degree of suppressibility of ACTH with exogenous glucocorticoid is determined. In classic Cushing disease due to a pituitary tumor, supraphysiologic doses of dexamethasone usually suppress the elevated plasma cortisol levels. Tumors responsible for the ectopic ACTH syndrome are classically unresponsive to these same doses of dexamethasone. High-dose dexamethasone suppression testing, as this diagnostic maneuver is called, is accomplished by (1) administering 2 mg of dexamethasone every 6 hours for 2 days and measuring urinary free cortisol or plasma cortisol on the second day, or (2) administering 8 mg of dexamethasone the night before obtaining an 8 AM plasma cortisol level. In both tests, the expected suppression of baseline urinary free cortisol and plasma cortisol should be 50% or greater if the Cushing syndrome is due to a pituitary adenoma (ie, Cushing disease). However, between 15% and 33% of patients with ectopic ACTH syndrome also meet these suppression criteria (false positives), mimicking Cushing disease. In addition, 10% to 25% of patients with Cushing disease fail to suppress with high-dose dexamethasone (false negatives). The overnight test probably has greater sensitivity and accuracy than the classic 2-day test, but neither test is considered useful any longer (see Chapter 9).

Additional provocative tests have been developed to improve the diagnostic discrimination between Cushing disease and ectopic ACTH syndrome using ovine CRH. Pituitary corticotrophs are normally responsive to CRH in Cushing disease and unresponsive when ectopic ACTH production or an adrenal lesion is responsible for the cortisol excess. A positive response to CRH is defined as a 50% or greater increase in plasma ACTH and a 20%

TABLE 21–2 Tumors responsible for the ectopic ACTH syndrome in two large contemporary series.

Tumor Type	Number of Patients (% of total) NIH Series (USA) (1983-2004)[a]	Number of Patients (% of total) St. Bartholomew's Hospital (UK) Cushing Disease Database (1969-2001)[b]
Carcinoid		
Thymic	5/90 (5.5%)	2/40 (5%)
Pulmonary	35/90 (39%)	12/40 (30%)
Appendiceal	1/90 (1%)	–
Pancreatic	1/90 (1%)	3/40 (7.5%)
Pulmonary tumorlets	1/90 (1%)	–
Neuroendocrine tumors	13/90 (14%)	2/40 (5%)
Gastrinoma	6/90 (6.7%)	–
Small cell lung cancer	3/90 (3.3%)	7/40 (17.5%)
Pheochromocytoma	5/90 (5.5%)	1/40 (2.5%)
Medullary thyroid cancer	2/90 (2.2%)	3/40 (7.5%)
Small cell colon	–	2/40 (5%)
Miscellaneous	1/90 (1%)	5/40 (12.5%)
Unknown	17/90 (18.9%)	5/40 (12.5%)

[a]Data from Ilias I, Torpy DJ, Pacak K, et al. Extensive clinical experience: Cushing's syndrome due to ectopic corticotropin secretion: twenty years' experience at the National Institute of Health. *J Clin Endocrinol Metab*. 2005;90:4955.

[b]Data from Isidori AM, Kaltsas GA, Pozza C. Extensive clinical experience: the ectopic adrenocorticotropin syndrome: clinical features, diagnosis, management, and long-term follow-up. *J Clin Endocrinol Metab*. 2006;91:371.

or greater increase in the plasma cortisol concentrations in the periphery. An increase in ACTH of 100% and in cortisol of more than 50% greatly reduces the likelihood of ectopic ACTH syndrome; however, false-positive and false-negative tests (up to 10%) have been reported. Moreover, in the rare instance of ectopic production of CRH (without concomitant ACTH) by a tumor, a false-positive result may be seen, leading to the erroneous diagnosis of pituitary-dependent Cushing disease.

For these reasons, most centers prefer inferior petrosal sinus sampling for ACTH before and after CRH administration, and this test is now the gold standard. The inferior petrosal sinuses drain the pituitary gland. Concomitant peripheral and petrosal sinus samples are obtained, and the central: peripheral ACTH ratio is calculated. In Cushing disease, the ratio should be greater than or equal to 2.0 in the basal state. After CRH administration, this ratio should be greater than or equal to 3.0 in pituitary-dependent Cushing disease. In the ectopic ACTH syndrome, the basal ratio is typically less than 2 and does not rise after the CRH. In rare instances of ectopic CRH syndrome, the basal ratio may be 2.0. The stimulation by CRH gives close to 100% discrimination between ectopic ACTH production and a pituitary tumor secreting ACTH. Generally, a combination of tests is performed to reach a biochemical diagnosis before extensive radiologic studies are undertaken.

The majority of patients (70% or more) with ectopic ACTH syndrome also cosecrete other hormones or tumor marker peptides, among them are carcinoembryonic antigen, somatostatin, calcitonin, gastrin, glucagon, vasoactive intestinal peptide (VIP), bombesin, pancreatic polypeptide, alpha-fetoprotein, and many others. The presence and secretion of these other hormones (in addition to ACTH) suggests that the source of ACTH is nonpituitary in these patients. Given the variety of peptides and the expense inherent in any screening paradigm, measuring these hormones in patients suspected of ectopic ACTH syndrome is not recommended.

The path to localize the tumor responsible for ectopic ACTH production generally starts with chest and abdominal imaging as most tumors are located there. Small cell carcinomas of the lung are usually visible on chest radiograph, whereas bronchial carcinoids are often difficult to detect on plain radiographs. In some situations, these tumors may require a long period (as many as 4 or 5 years) of close follow-up before the tumors are detected. Chest computed tomography (CT) scanning should be employed in all subjects with ectopic ACTH to rule out a lung or mediastinal lesion (such as a thymic carcinoid). Abdominal CT scanning is also performed in these patients to confirm the presence of bilateral adrenal enlargement, a sine qua non of the ectopic ACTH syndrome, and to screen for other possible abdominal tumors responsible for the syndrome (eg, pheochromocytoma, islet cell tumor). In the radiologic evaluation of Cushing syndrome, it is always important to bear in mind that the presence of a pituitary microadenoma on magnetic resonance imaging (MRI) does very little to support the diagnosis of pituitary-dependent Cushing disease—as opposed to an ectopic tumor producing ACTH—because of the great numbers (10%-20%) of

normal individuals with incidental pituitary microadenomas (see Chapter 4). An exception is the pituitary tumor greater than 6 mm. If a lesion of that size is seen, then a pituitary etiology can be strongly considered, if the remainder of the diagnostic testing supports a central etiology.

Octreotide scanning, another important diagnostic technique, can successfully localize tumors responsible for ectopic ACTH production because up to 80% of such tumors express somatostatin receptors. Iodinated or, more recently, indium-111(^{111}In)-labeled octreotide scanning has demonstrated medullary carcinomas of the thyroid, small cell lung cancers, islet cell tumors, pheochromocytomas, and other tumors.

The ideal imaging algorithm for occult tumors causing ectopic ACTH syndrome is controversial. Functional imaging with higher than standard doses of the somatostatin analog [^{111}In]-diethylene triamine pentaacetate-D-Phe-pentetreotide (OctreoScan) or with [^{18}F]-fluorodeoxyglucose positron emission tomography (FDG-PET) was prospectively evaluated in 17 patients with the ectopic ACTH syndrome (based on inferior petrosal sinus sampling) in whom standard imaging studies were inconclusive. Nieman and colleagues found that FDG-PET did not uncover tumors that were not seen on CT or MRI scans. These investigators, however, found OctreoScans to be useful (enhanced sensitivity) in combination with CT/MRI imaging and advocated this diagnostic approach. The somatostatin receptor PET tracer ^{68}Ga-DOTA-NOC (^{68}Ga-labeled [1,4,7,10-tetraazacyclododecane acid]-1-Nal3-octreotide) PET/CT scanning has also been compared to contrast-enhanced CT for localizing tumors in ectopic ACTH, and while sensitivity may be lower, false-positive rates are very low.

Clinical Features

Cushing syndrome causes truncal obesity, violaceous striae, hypertension, fatigue, glucose intolerance, osteopenia, muscle weakness, moon facies, easy bruisability, buffalo hump, depression, hirsutism, and edema. Patients with ectopic ACTH syndrome may show some, all, or none of these features depending on the underlying tumor. It has been appreciated from the initial descriptions of this syndrome that these patients typically present with myopathy, weight loss, and electrolyte and metabolic disturbances more commonly than with the classic features of slowly developing Cushing disease. Hyperpigmentation is also recognized as more common in the ectopic ACTH syndrome than in Cushing disease. Cortisol excess in older men, especially those at risk for lung tumors, is most commonly due to ectopic ACTH production, whereas ACTH-producing pituitary tumors predominate in young and middle-aged women. Glucose intolerance or frank diabetes and hypokalemic alkalosis are typical metabolic disturbances of the ectopic ACTH syndrome. Because of the extreme elevation in plasma cortisol levels in many of these patients, they are at considerable risk for and often succumb to overwhelming opportunistic infections, often with fungal pathogens.

A critical caveat to remember in the clinical evaluation of patients with ACTH-dependent Cushing disease is that slowly

growing and occult tumors producing ACTH may present in exactly the same way as classic Cushing disease due to a pituitary tumor. Therefore, both the clinical findings and the laboratory studies summarized earlier show considerable overlap and may engender confusion in distinguishing these occult tumors from a pituitary lesion.

A small number of patients who have classic features of Cushing syndrome have been shown to have abnormal expression of receptors in their adrenal glands as the cause of their hypercortisolism. The pathophysiology of this form of Cushing syndrome is ACTH-independent, because other hormones are driving the glucocorticoid hypersecretion. This has been likened by Lacroix and colleagues to a *gain of function* mutation causing constitutive activation of a G protein–coupled receptor. Ectopic expression of receptors for gastric inhibitory peptide (GIP), vasopressin (V_2 and V_3), serotonin (5-HT_7), and beta adrenergic agonists have been reported. Altered activity or greater expression of eutopic receptors for serotonin (5-HT_4), LH, and vasopressin (V_1) in adrenal tissue can also cause cortisol excess. In the case of GIP, food-stimulated cortisol hypersecretion has been described. In a case report of excessive LH receptor expression in the adrenals associated with macronodular adrenal hyperplasia, the patient had mild Cushingoid features with pregnancy and the gradual development of full-blown Cushing syndrome with menopause. It is noteworthy that many patients with ectopic or eutopic receptor-mediated Cushing syndrome have macronodular adrenal hyperplasia.

Treatment

Because of the morbidity and mortality of untreated and often severe hypercortisolism (especially due to ectopic ACTH syndrome), therapy is often rapidly instituted to reduce complications. Surgical excision and debulking of the tumor, along with chemotherapy and/or radiation, offer the chance for immediate relief, if the resection is complete. If the tumor cannot be removed, then medical therapy with adrenal steroidogenesis blockers, either alone or in combination, may be instituted including metyrapone, ketoconazole, and mitotane. Limited success is reported with dopamine agonist therapy with cabergoline or with somatostatin receptor-based therapy with octreotide. The glucocorticoid, progesterone, and androgen receptor antagonist mifepristone has also been tested in patients with ectopic ACTH syndrome. The agent may work quickly to block glucocorticoid action and is approved for control of hyperglycemia in patients with hypercortisolism secondary to endogenous Cushing syndrome who have failed surgery or are not candidates for it. The anesthetic agent etomidate can be used to turn off steroidogenesis rapidly in the acutely ill patient deteriorating rapidly from consequences of severe hypercortisolism. All medical therapies require careful dosing and close monitoring for toxicity and for complete adrenal blockade with the emergence of adrenal insufficiency. To gain definitive control of hypercortisolism, bilateral adrenalectomy is an additional option. Expert care by an experienced team of endocrinologists, oncologists, and radiologists is essential for successfully treating these complex and ill patients.

SYNDROME OF INAPPROPRIATE ANTIDIURETIC HORMONE SECRETION

The syndrome of inappropriate antidiuretic hormone (SIADH) secretion is characterized by inappropriate retention of water such that hypertonic or inappropriately concentrated urine is excreted in the presence of euvolemic hyponatremia. There are many well-recognized causes of this syndrome (Table 21–3).

Etiology and Pathogenesis

Tumors are a common cause of SIADH (see Table 21–3). Bronchogenic carcinoma, particularly small cell carcinoma, has been associated with this syndrome since its initial description in 1957 by Schwartz and Bartter. Small cell carcinoma accounts for 80% of lung tumors associated with SIADH, but only 3% to 15% of patients with this tumor have SIADH. Most of these tumors, even from patients without the clinical syndrome, contain ADH by immunostaining. Other tumors that cause the syndrome include breast, pancreatic, and thymic carcinomas in addition to those listed in Table 21–3. Other causes of excessive ADH release and

TABLE 21–3 Causes of SIADH.

Tumors
    Small cell carcinoma of lung
    Squamous cell carcinoma of lung
    Cancers of the head and neck
    Carcinoma of the duodenum, pancreas, ureter, prostate, uterus, and nasopharynx
    Mesothelioma
    Thymoma
    Hodgkin lymphoma
    Leukemia

Central nervous system disorders
    Brain tumors (primary and metastatic)
    Brain abscess
    Subdural hematoma, subarachnoid hemorrhage
    Meningitis, encephalitis
    Systemic lupus erythematosus
    Demyelinating disorders
    Head trauma
    Acute intermittent porphyria
    Guillian-Barré syndrome
    Spinal cord lesions
    Pituitary stalk lesion
    Transsphenoidal adenectomy
    Hydrocephalus

Drugs
    Nicotine, phenothiazines, tricyclic antidepressants, nonsteroidal anti-inflammatory agents, cyclophosphamide, vincristine, chlorpropamide, colchicine, selective serotonin reuptake inhibitors (sertraline, fluoxetine), azithromycin, thiazide diuretics

Pulmonary disorders
    Tuberculosis
    Fungal, bacterial, viral, mycoplasmal pneumonia
    Empyema, lung abscess
    Chronic obstructive pulmonary disease

Nephrogenic syndrome of inappropriate antidiuresis

Data from Verbalis JG. Disorders of body water homeostasis. *Best Pract Res Clin Endocrinol Metab.* 2003 Dec;17(4):471-503.

of hyponatremia in patients with cancer such as anticancer therapy, other drug treatments, pain, nausea, and vomiting should be considered in this setting.

Excessive production of ADH by tumors leads to an inability to excrete free water. This is due to the fact that tumors release ADH independent of serum osmolality. In addition to the production of ADH, many tumors also contain the mRNA for the hormone atrial natriuretic peptide (ANP), which likely causes no clinical symptoms.

Clinical and Laboratory Features

SIADH is the most common cause of hyponatremia in hospitalized patients. It may present with symptoms due to water intoxication and hyponatremia. Although many patients are asymptomatic, depending on the magnitude and chronicity of their hyponatremia, symptomatic individuals usually have fatigue, headache, nausea, and anorexia initially, which can progress to altered mental status, gait instability and falls, seizures, coma, and even death. Most patients experience weight gain due to water retention but do not have edema. Significant clinical symptoms usually do not develop unless the serum sodium is 125 mEq/L or less, and there is usually a correlation between the level of symptomatology in these patients and their serum sodium values.

Patients with SIADH exhibit hyponatremia, serum hypo-osmolality, less than maximally dilute urine, and the presence of sodium in the urine. Clinically, the diagnosis of SIADH cannot be made unless there is euvolemia with intact renal, adrenal, and thyroid function. Cirrhosis with ascites, nephrosis, congestive heart failure, and recent diuretic use must be excluded. Urinary sodium levels are usually high (>30 mmol/L), and serum urea nitrogen levels are typically low, as are serum uric acid levels. Rarely is it necessary or even helpful to measure ADH levels to make this diagnosis. Although these determinations are available, they are even sometimes misleading. Recent studies support the possibility that copeptin, the C-terminal part of ADH molecule, may be a more stable surrogate marker for ADH and may prove useful ultimately for diagnosis. Once the diagnosis of SIADH is made, all possible causes should be considered (see Table 21–3).

Water restriction and demeclocycline are the traditional therapies. Studies support the efficacy of selective vasopressin antagonists (vaptans) in the treatment of SIADH, and this approach may be considered in cases refractory to the above measures and/or before the use of demeclocycline. In acutely symptomatic patients or when the serum sodium is dangerously low and symptoms of hyponatremia are present, infusion of hypertonic saline and administration of loop diuretics (eg, furosemide) are the treatments of choice. Many patients with SIADH due to neoplasms improve and even remit with effective therapy for their underlying cancer.

NON-ISLET CELL TUMOR-INDUCED HYPOGLYCEMIA

Tumor-induced hypoglycemia is rare and most commonly caused by eutopic insulin production by pancreatic islet cell tumors (insulinomas). The remainder of cases (20%) are primarily due to tumor-derived insulin-like growth factor-2 (IGF-2). Ectopic insulin secretion by tumors has only rarely been reported, and other tumoral mechanisms inducing endogenous insulin are also rare (tumor production of GLP-1 or antibodies against insulin or its receptor).

The most common sources of ectopic IGF-2 driving non-islet cell tumor-induced hypoglycemia are either: (1) large (>2 kg), slow-growing and well-differentiated solitary fibrous tumors of mesenchymal origin (22%) that primarily arise in the pleura but also occur at other sites (bladder, uterus); or (2) hepatocellular carcinomas (17%). Other less commonly reported tumors associated with IGF-2-induced hypoglycemia include adenocarcinomas, gastrointestinal stromal tumors, sarcomas, adrenocortical carcinoma, phyllodes breast tumors, and renal cell carcinomas. Subtle signs of fasting hypoglycemia may be the presenting symptom in these cases, or hypoglycemia may be detected during the course of tumor treatment. As discussed fully in Chapter 18, signs and symptoms of fasting hypoglycemia include sweating, intense hunger, anxiety, altered consciousness, and visual and behavioral changes.

Non-islet cell tumor, IGF-induced hypoglycemia, first described in 1930 by Doege, was identified in a patient with of hypoglycemia associated with a pleural mesenchymal tumor (now specifically known as the Doege-Potter syndrome). This was originally attributed to glucose utilization by these large tumors. However, hypoglycemia is only seen in 5% of such tumors. The actual pathogenesis for non-islet cell tumor-induced hypoglycemia by these and other tumors was elucidated in the last few decades in parallel with our increased understanding of IGF biology. *IGF-2* gene expression is increased in these tumors; however, posttranslational processing is disrupted such that the majority of circulating IGF-2 is a precursor pro-IGF-2 protein (big IGF-2). Big IGF-2 is less likely to form normal ternary complexes with IGF-binding protein 3 (IGF-BP 3) and an acid labile subunit, and therefore is more bioavailable. In addition, it is postulated that big IGF-2 may displace normal IGF-1 and IGF-2 from IGF-BP, as free levels of these hormones are also increased in some patients. In addition, big IGF-2 is thought to feedback inhibit growth hormone (GH) secretion, thus resulting in even lower levels of IGF-BP. This vicious cycle enhances the bioavailability of free IGF peptides whose affinity for the insulin receptor is only tenfold lower than that of insulin itself. The typical hormonal profile of patients with this syndrome includes suppressed levels of insulin, C-peptide, and total IGF-1; and normal or only slightly elevated levels of total IGF-2, while levels of free IGF-2 are markedly (20-fold) increased.

Stimulated glucose disposal in skeletal muscle by IGF binding to the insulin receptor is thought to be a major contributor to hypoglycemia in this syndrome. Diminished hepatic glucose production due to hepatic insulin receptor–mediated IGF-2 effects may also contribute. While GH levels are suppressed, some acromegalic features, thought to be secondary to IGF-2 binding to the type 1 IGF receptor, can also be present (skin tags, rhinophyma).

Treatment of this paraneoplastic syndrome usually involves curative resection or debulking of the tumor. Preoperatively, patients often require continuous glucose infusions to control

their symptoms, and glucagon can be used acutely to raise blood glucose levels. If the lesion is benign, surgery usually brings relief of the hypoglycemia or even definitive cure. In the absence of a surgical cure, additional palliative measures to control hypoglycemic symptoms have been used with some success, including (1) glucocorticoid treatment to suppress tumor production of IGF-1 and induce hepatic gluconeogenesis, and/or (2) GH treatment to increase serum levels of IGF-BP, as well as to induce hepatic gluconeogenesis.

OTHER HORMONES SECRETED BY TUMORS

1. GROWTH HORMONE–RELEASING HORMONE AND GROWTH HORMONE

Acromegaly is primarily due to eutopic production of GH by pituitary adenomas. The remainder of cases (<2%) are due to ectopic production of growth hormone releasing hormone (GHRH) by well-differentiated NETs, the majority of which are bronchial carcinoids or pancreatic endocrine tumors. Ectopic GH overproduction is extremely rare and primarily due to tumors in the sphenoid sinus which may either be pituitary remnants in the embryological pathway of Rathke's pouch invagination and/or inferior protrusions of pituitary tumors due to sellar defects.

Ectopic secretion of GHRH was first described in 1974. Indeed, this hypothalamic hormone was first isolated a decade later from patients with pancreatic GHRH-producing tumors. Excessive peripheral production of GHRH causes hyperplasia of somatotrophs in the pituitary. Resultant elevated levels of circulating GH and IGF-1, lack of suppression of GH by an oral glucose load, and clinical features of acromegaly in this setting are indistinguishable from those caused by GH-secreting pituitary adenomas (see Chapter 4), which can lead to misdiagnosis and unnecessary transsphenoidal surgery. Ectopic GHRH as a cause of acromegaly can be simply and specifically diagnosed by documentation of elevated circulating GHRH levels. However, because ectopic GHRH is rarely a cause of acromegaly, this assay is best reserved for atypical presentations associated with additional signs and symptoms specific to those tumors that are the most frequent source of ectopic GHRH (carcinoid and pancreatic tumors). It should also be noted that in rare instances, hypothalamic GHRH-secreting tumors (hamartomas, gliomas, and gangliocytomas) can also produce excessive GHRH and acromegaly. Because the site of GHRH production is within the hypothalamus, these cases do not represent ectopic GHRH secretion. Moreover, from a diagnostic standpoint, circulating GHRH levels in these cases are not elevated.

Carcinoid tumors, primarily in the lung, but occasionally at other sites (gastrointestinal tract or thymus) are the most frequent peripheral source of GHRH (>50%), followed by pancreatic islet cell tumors (30%). Other NETs associated with GHRH-induced acromegaly include pheochromocytoma, small cell lung cancer, adrenal adenoma, medullary thyroid carcinoma, endometrial cancer, and breast cancer. Interestingly, while 50% of carcinoid

tumors and 25% of endocrine tumors are immunoreactive for GHRH, a minority of these tumors is associated with elevated circulating levels of bioactive GHRH. Slow-growing GHRH-producing carcinoid tumors, the majority of which are bronchial in origin, are responsible for an insidious presentation of acromegaly that is often associated with additional symptoms specific to the tumor (eg, wheezing). While these tumors can be easily detected by routine imaging, octreotide scanning may also be helpful as somatostatin receptors are often present in GHRH-secreting tumors. While 50% of GHRH-secreting tumors have metastasized at time of diagnosis, long-term survival is high, consistent with the indolent nature of many NETs. In addition to surgical resection, somatostatin analogs, particularly those binding to multiple somatostatin receptor subtypes, can be useful in the management of ectopic GHRH-induced acromegaly.

2. CALCITONIN

Circulating levels of calcitonin, formed normally by cleavage of procalcitonin within the C cells of the thyroid in response to hypercalcemia, are elevated when overproduced eutopically by medullary thyroid cancer (see Chapters 7 and 24). NETs of the foregut (primarily pancreas and lung) are the primary source of ectopic calcitonin secretion. Consistent with our lack of knowledge regarding the precise physiologic role of calcitonin, there is often no clinical syndrome associated with its ectopic production. However, after an initial diagnosis, calcitonin levels have been used as a tumor marker to follow responses to treatment. While the incidence of calcitonin hypersecretion appears to be higher in lung (53%) versus pancreatic NETs (11%), pancreatic NETs appear to be the most common source of ectopic calcitonin due to their higher prevalence in reported series. In almost all reported cases of calcitonin-secreting NETs, liver or bone metastases are present. Calcitonin hypersecretion by NETs (or medullary thyroid carcinoma) is often associated with diarrhea (30% of cases) that is poorly responsive to somatostatin, and is likely attributable to calcitonin and/or other gut derived peptides frequently cosecreted by NETs. In the absence of renal failure, which is also associated with elevated calcitonin levels, or evidence of a thyroid tumor, high serum calcitonin levels should suggest a possible foregut NET source. It should be noted that while circulating procalcitonin levels are markedly increased in response to certain inflammatory stimuli, such as bacteremia, calcitonin levels are normal or only moderately elevated in these settings.

3. GONADOTROPINS

The gonadotropins, luteinizing hormone (LH), follicle-stimulating hormone (FSH), and human chorionic gonadotropin (hCG), are structurally related heterodimeric ($\alpha\beta$) glycoproteins that share a common alpha subunit. While hCG is primarily known for its role as a placenta hormone, being produced by trophoblastic cells during pregnancy, all three are in fact produced by the normal pituitary. While LH and FSH are rarely produced ectopically, hCG overproduction by a variety of tumors is common. Ectopic hCG production must be distinguished from eutopic secretion

by trophoblastic tumors (hydatidiform mole, choriocarcinoma). Trophoblastic tumors frequently secrete intact hetrodimers of hCG that are either normally or heavily (hyper-) glycosylated, with the later moiety being particularly associated with choriocarcinoma and normal pregnancy implantation (extravillous invasion of cytotrophoplasts). Germ cell tumors of the testes or ovary also secrete intact hCG (or beta subunits); indeed, serum levels of intact hCG, which are secreted by 20% to 50% of testicular germ cell tumors, are used for diagnosis and monitoring. In contrast, non-trophoblastic, non-germ cell tumors primarily secrete the free beta subunit of hCG (hCGβ), which has been demonstrated to promote tumor growth by inhibiting apoptosis. While hCG binds to a common receptor with LH (the LH/CG receptor), due to the high homology of their beta subunits, which arose from a gene duplication event, the means by which hCGβ exerts its biological effects is not well understood. High circulating levels of hCGβ occur commonly with a large number of tumor types, including: genitourinary tumors (10%-75% of bladder cancers, 25% of renal cancers), non-trophoblastic gynecologic cancers (20%-75% of endometrical cancers), gastrointestinal cancers (40%-60% of biliary, pancreatic and gastric cancers; 15%-20% of liver and colorectal cancers), and lung cancers (13% of small cell cancers, 6% of all cancers). In all cases, increased serum concentrations of hCG are associated with poorer prognosis. Increases in serum immunoreactive hCG, while mild, have frequently been reported in breast cancer (10%-50%) but are of unclear significance, in part due to their possible attribution to the normal increase in pituitary hCG secretion that occurs with age and menopausal status. Given the frequent adverse association between tumor hCG production and prognosis, hCG has been a target of interest for cancer therapeutics. Depending on the hCG isoform produced (intact heterodimer vs beta subunit) and the age and gender of the patient, the clinical signs and symptoms of hCG excess are variable (precocious puberty in children, uterine bleeding in women, hypogonadism in men) and may be absent.

ONCOGENIC OSTEOMALACIA

Etiology and Clinical Features

Oncogenic osteomalacia is a syndrome associated with unusual mesenchymal tumors and rarely with prostate cancer and multiple myeloma. Patients with the syndrome have hypophosphatemia, renal phosphate wasting, and typically low or inappropriately normal serum levels of 1,25-(OH)$_2$ D. Alkaline phosphatase activity, reflecting bone turnover, is often elevated. Levels of calcium are typically normal, although in some cases PTH levels are elevated and calcium levels low. Hypophosphatemia is due to reduced renal phosphate reabsorption. Clinical symptoms include bone pain, proximal muscle weakness, fractures, back pain, waddling gait, and progressive debility. The syndrome often poses a significant diagnostic dilemma to clinicians, because the tumors responsible for it may be very small, obscurely situated, and difficult to identify. Phosphate depletion and low 1,25-(OH)$_2$ D levels lead to poor bone mineralization and osteomalacia. Typically, the diagnosis of osteomalacia is not suspected for years despite the presence of classic symptoms. The humoral basis for the syndrome is supported by the observation that the biochemical abnormalities remit and the osteomalacia heals when the tumor responsible for it is removed.

Pathology and Pathogenesis

Tumors responsible for oncogenic osteomalacia are usually small and grow slowly. The histology is described as phosphaturic mesenchymal tumor mixed connective tissue variant. Because of the unusual histology and often unanticipated locations for these tumors, the syndrome has been dubbed "strange tumors in strange places" by Weiss and colleagues. The range of locations for these tumors includes the lower extremities (45%), head and neck (27%), and upper extremities (17%). In a review of head and neck tumors that cause oncogenic osteomalacia, Gonzalez-Compta and coworkers noted that in 57% and 20% of cases, respectively, tumors were in the sinonasal and mandibular areas. These tumors are generally solitary lesions and benign, although multifocal lesions have been reported as well as malignant phenotypes.

The histologic spectrum of tumors includes hemangiomas, hemangiopericytomas, angiosarcomas, chondrosarcomas, prostate cancer (particularly castrate-resistant), schwannomas, neuroendocrine lesions, mesenchymal tumors, and uncommonly MM. Many of these tumors, classified pathologically as mixed connective tissue tumors, are located in bone. Osteoclast-like giant cells and stromal cells as well as highly vascular features characterize these tumors. Delays in diagnosis of as long as 19 years have been reported, thereby supporting the hypothesis of slow growth of these tumors.

Most of the tumors responsible for this syndrome overexpress fibroblast growth factor (FGF) 23, a protein important in normal phosphate homeostasis (see Chapter 8). FGF23 overexpression appears to be responsible for key aspects of oncogenic osteomalacia, and levels of FGF23 are often dramatically elevated. These levels fall in response to tumor resection, and the biochemical changes and osteomalacia typically resolve after surgery. In addition to FGF23, other genes have been implicated in the pathogenesis of oncogenic osteomalacia including FGF7, the matrix extracellular phosphoglycoprotein (MEPE), and frizzled related protein 4 (FRP4).

Depending on the series, approximately 40% to 60% of FGF23-producing tumors express a fibronectin-FGF-receptor 1 (FN1-FGFR1) fusion gene and some even an FN1-FGF1 fusion gene at a lower prevalence. These findings may be of pathogenetic significance in that FN signaling can increase FGF23 production through an autocrine/paracrine mechanism. Therapies targeting the FGFR1 are being developed as medical therapies for unresectable and metastatic lesions in patients with oncogenic osteomalacia along with antibodies directed against FGF23 itself.

Localization

Tumors causing oncogenic osteomalacia are often small and difficult to find. Thorough cranial, chest, abdominal, pelvic, and extremity imaging are essential for anatomic localization toward

the goal of surgical cure. Somatostatin receptor-based scintigraphy is often an early diagnostic step using 111-Indium octreotide SPECT/CT. 68-Gallium DOTA-TATE PET/CT has also been successful in finding these tumors, which should then be definitively localized by MRI scanning. For tumors that resist localization by these means, selective venous sampling for intact FGF23 has been successfully employed.

Comparison with Other Disorders of FGF23 Overproduction

Tumor-induced osteomalacia and X-linked hypophosphatemic rickets share many features. These features include rickets in growing children and osteomalacia in adults along with hypophosphatemia, low or inappropriately normal $1,25\text{-}(OH)_2$ D, and elevated FGF23 levels. Despite these similarities, there are several unresolved differences between the two syndromes. One is that levels of $1,25\text{-}(OH)_2$ D are inappropriately normal in patients with X-linked hypophosphatemic rickets and frankly low in oncogenic osteomalacia. In addition, patients with X-linked hypophosphatemic rickets demonstrate osteosclerosis and enthesopathy (calcification of tendons and ligaments), which are not seen in oncogenic osteomalacia.

X-linked hypophosphatemic rickets is due to defective functioning of the *PHEX* gene product (PHEX), a protein that is homologous to neutral, membrane-bound endopeptidases. PHEX is thought to activate or inactivate a circulating factor involved in phosphate metabolism, termed phosphatonin. It has long been proposed that the normal function of phosphatonin was to block renal phosphate reabsorption. FGF23 has been shown to inhibit phosphate uptake in kidney cells. PHEX, the endopeptidase product of the *PHEX* gene, can degrade FGF23, but FGF23 is not thought to be a substrate *in vivo* for PHEX. There remain several crucial details to establish regarding the pathogenesis and manifestations of oncogenic osteomalacia and X-linked hypophosphatemic rickets.

GUT HORMONES

Gut hormones, including VIP, somatostatin, gastrin-releasing hormone, and pancreatic polypeptide can be produced ectopically, particularly by NETs (carcinoid, pancreatic islet, and small cell lung cancer). Of these, only ectopic VIP secretion produces a clinical syndrome of voluminous secretory diarrhea, achlorhydria, and hypokalemia. This constellation of symptoms (Verner-Morrison syndrome) has been termed pancreatic cholera when caused by a pancreatic VIP-secreting islet cell tumor. A variety of nonendocrine and neuroendocrine tumors can express and release VIP, including small cell carcinomas of the lung, carcinoids, pheochromocytoma, medullary carcinomas of the thyroid, and some colonic adenocarcinomas. VIP and its closely related hormone PACAP (pituitary adenylate cyclase activating peptide) both have effects on the growth and angiogenesis of the tumors that produce them. These peptides and their receptors are being targeted for the development of new therapies including radioanalogues that may prove useful for molecular localization of the tumors and their metastatic deposits.

REFERENCES

Hypercalcemia of Malignancy

Martin TJ. Parathyroid hormone-related protein, its regulation of cartilage and bone development, and role in treating bone disease. *Physiol Rev*. 2016;96:831. [PMID: 27142453]

Rizzoli R, Body JJ, Brandi ML, et al. Cancer-associated bone disease. *Osteoporos Int*. 2013;24:2929. [PMID: 24146095]

Roodman GD. Genes associate with abnormal bone cell activity in bone metastasis. *Cancer Metastasis Rev*. 2012;31:569. [PMID: 22706844]

Sternlicht H, Glezerman IG. Hypercalcemia of malignancy and new treatment options. *Ther Clin Risk Manag*. 2015;11:1779. [PMID: 26675713]

Wagner J, Arora S. Oncologic metabolic emergencies. *Emerg Med Clin North Am*. 2014;32:509. [PMID: 25060247]

Syndrome of Inappropriate Antidiuretic Hormone Secretion

Berardi R, Rinaldi S, Caramanti S, et al. Hyponatremia in cancer patients: time for a new approach. *Crit Rev Hemat Oncol*. 2016;102:15. [PMID: 27066939]

Fenske W, Sandner B, Christ-Crain M. A copeptin–based classification of the osmoregulatory defects in the syndrome of inappropriate antidiuresis. *Best Pract Res Clin Endo Metab*. 2016;30:219. [PMID: 27156760]

Grant P, Ayuk J, Bouloux PM, et al. The diagnosis and management of inpatient hyponatraemia and SIADH. *Eur J Clin Invest*. 2015;45:888. [PMID: 25995119]

Peri A. The use of vaptans in clinical endocrinology. *J Clin Endo Metab*. 2013;90:1321. [PMID: 23401044]

Verbalis JG, Goldsmith SR, Greenberg A, et al. Diagnosis, evaluation and treatment of hyponatremia: expert panel recommendations. *Am J Med*. 2013;126:S1. [PMID: 24074529]

Ectopic ACTH Syndrome

Alexandraki K, Grossman AB. The ectopic ACTH syndrome. *Rev Endocr Metab Disord*. 2010;11:117. [PMID: 20544290]

Alexandraki K, Grossman AB. Therapeutic strategies for the treatment of severe Cushing's syndrome. *Drugs*. 2016;76:447. [PMID: 26833215]

Daniel E, Aylwin S, Mustafa O, et al. Effectiveness of metyrapone in treating Cushing's Syndrome: a retrospective multicenter study in 195 patients. *J Clin Endocrinol Metab*. 2015;100:4146. [PMID: 26353009]

Goroshi MR, Jadhav SS, Lila AR, et al. Comparison of 68Ga-DOTANOC PET/CT and contrast-enhanced CT in localization of tumors in ectopic ACTH syndrome. *Endocr Connect*. 2016;5:83. [PMID: 27006371]

Ilias I, Torpy DJ, Pacak K, et al. Cushing's syndrome due to ectopic corticotropin secretion: twenty years' experience at the National Institutes of Health. *J Clin Endocrinol Metab*. 2005;90:4955. [PMID: 15914534]

Lacroix A, Baldacchino V, Bourdeau I, Hamet P, Tremblay J. Cushing's syndrome variants secondary to aberrant hormone receptors. *Trends Endocrinol Metab*. 2004;15:375. [PMID: 15380809]

Lacroix A, Feelders RA, Stratakis CA, Nieman LK. Cushing's syndrome. *Lancet*. 2015;386:913. [PMID: 26004339]

Pozza C, Graziadio C, Giannetta E, Lenzi A, Isidori AM. Management strategies for aggressive Cushing's Syndrome: from macroadenomas to ectopics. *J Oncol*. 2012;685213. [PMID: 22934113]

Yogi-Morren D, Habra MA, Faiman C, et al. Pituitary MRI findings in patients with pituitary and ectopic ACTH-dependent Cushing Syndrome: does a 6-mm pituitary tumor size cut-off value exclude ectopic ACTH syndrome? *Endocr Pract*. 2015;10:1098. [PMID: 26121435]

Witek P, Witek J, Zielinski G, et al. Ectopic Cushing's syndrome in light of modern diagnostic techniques and treatment options. *Neuroendo Lett*. 2015;36:201. [PMID: 26313384]

Hypoglycemia

Bodnar TW, Acevedo MJ, Pietropaolo M. Management of non islet-cell tumor hypoglycemia: a clinical review. *J Clin Endocrinol Metab*. 2014;99:713. [PMID: 24423303]

Gonadotropins

Cole LA. hCG, five independent molecules. *Clin Chim Acta*. 2012;413:48. [PMID: 22027338]

Sahoo S, Singh P, Kalha B, Singh O, Pal R. Gonadotropin-mediated chemoresistance: delineation of molecular pathways and targets. *BMC Cancer*. 2015;15:931. [PMID: 26608647]

Stenman UH, Alfthan H, Hotakainen K. Human chorionic gonadotropin in cancer. *Clin Biochem*. 2004;37:549. [PMID: 15234236]

GHRH and Growth Hormone

Borson-Chazot F, Garby L, Raverot G, et al; GTE group. Acromegaly induced by ectopic secretion of GHRH: a review 30 years after GHRH discovery. *Ann Endocrinol (Paris)*. 2012;73:497. [PMID: 23122576]

Ghazi AA, Amirbaigloo A, Dezfooli AA, et al. Ectopic acromegaly due to growth hormone releasing hormone. *Endocrine*. 2013;43:293. [PMID: 22983831]

Ramírez C1, Hernández-Ramírez LC, Espinosa-de-los-Monteros AL, et al. Ectopic acromegaly due to a GH-secreting pituitary adenoma in the sphenoid sinus: a case report and review of the literature. *BMC Res Notes*. 2013;6:411. [PMID: 24119925]

Calcitonin

Nozières C, Chardon L, Goichot B, et al. Neuroendocrine tumors producing calcitonin: characteristics, prognosis and potential interest of calcitonin monitoring during follow-up. *Eur J Endocrinol*. 2016;174:335. [PMID: 26671974]

Sedef AM, Kose F, Mertsoylu H, Ozyilkan O. Procalcitonin as a biomarker for infection-related mortality in cancer patients. *Curr Opin Support Palliat Care*. 2015;9:168. [PMID: 25872114]

Wuilmet L, Jovenin N, Larbre H, et al. Digestive calcitonin-secreting tumors of the foregut: comparison with non-calcitonin-secreting tumors. *Eur J Gastroenterol Hepatol*. 2006;18:951. [PMID: 16894307]

Oncogenic Osteomalacia

Bansal S, Khazim K, Suri R, et al. Tumor induced osteomalacia: associated with elevated circulating levels of fibroblast growth factor-7 in additon to fibroblast growth factor-23. *Clin Nephrol*. 2016;85:57. [PMID: 26521888]

Breer S, Brunkhorst T, Beil FT, et al. 68Ga DOTA-TATE PET/CT allows tumor localization in patients with tumor-induced osteomalacia but negative 111In-octreotide SPECT/CT. *Bone*. 2014;64:222. [PMID: 24769333]

Chong WH, Andreopoulou P, Chen CC, et al. Tumor localization and biochemical response to cure in tumor-induced osteomalacia. *J Bone Miner Res*. 2013;28:1386. [PMID: 23362135]

Endo I, Fukumoto S, Ozono K, et al. Nationwide survey of fibroblast growth factor 23 (FGF23)-related hypophosphatemic diseases in Japan: prevalence, biochemical data and treatment. *Endocr J*. 2015;62:811. [PMID: 26135520]

Imanishi Y, Hashimoto J, Ando W, et al. Matrix extracellular phosphoglycoprotein is expressed in causative tumors of oncogenic osteomalacia. *J Bone Miner Metab*. 2012;30:93. [PMID: 21739089]

Jiang Y, Xia WB, Xing XP, et al. Tumor-induced osteomalacia: an important cause of adult-onset hypophosphatemic osteomalacia in China: report of 39 cases and review of the literature. *J Bone Miner Res*. 2012;27:1967. [PMID: 22532501]

Lee JC, Su SY, Changou CA, et al. Characterization of FN1-FGFR1 and novel FN1-FGF1 fusion genes in a large series of phosphaturic mesenchymal tumors. *Mod Pathol*. 2016;29:1335. [PMID: 27443518]

Tajima S, Fukayama M. Fibroblast growth factor receptor 1 (FGFR1) expression in phosphaturic mesenchymal tumors. *Int J Clin Exp Pathol*. 2015;8:9422. [PMID: 26464698]

Weiss D, Bar RS, Weidner N, Wener M, Lee F. Oncogenic osteomalacia: strange tumours in strange places. *Postgrad Med J*. 1985;61:349. [PMID: 4022870]

Zhang J, Zhu Z, Zhong D, et al. 68Ga DOTATATE PET/CT is an accurate imaging modality in the detection of cultprit tumors causing osteomalacia. *Clin Nucl Med*. 2015;40:642. [PMID: 26053726]

Gut Hormones

Bodei L, Ferone D, Grana CM, et al. Peptide receptor therapies in neuroendocrine tumors. *J Endocrinol Invest*. 2009;32:360. [PMID: 19636207]

Moody TW, Nuche-Berenguer B, Jensen RT. Vasoactive intestinal peptide/pituitary adenylate cyclase activating polypeptide, and their receptors and cancer. *Curr Opin Endocrinol Diabetes Obes*. 2016;23:38. [PMID: 26702849]

Tang B, Yong X, Xie R, Li QW, Yang SM. Vasoactive intestinal peptide receptor-based imaging and treatment of tumors (Review). *Int J Oncol*. 2014;44:1023. [PMID: 24481544]

This page intentionally left blank

22

Multiple Endocrine Neoplasia

David G. Gardner, MD, MS

ACTH	Adrenocorticotropic hormone	**MEN**	Multiple endocrine neoplasia
CT	Computed tomography	**MRI**	Magnetic resonance imaging
GDNF	Glial cell line-derived neurotrophic factor	**PCR**	Polymerase chain reaction
GDNFR	Glial cell line-derived neurotrophic factor receptor	**PTH**	Parathyroid hormone
		RET	Rearranged during transfection proto-oncogene
GNAS	Gene encoding the alpha subunit of a stimulatory G protein	**VIP**	Vasoactive intestinal peptide
HLXB9	Homeobox protein HB9	**ZES**	Zollinger-Ellison syndrome
MCT	Medullary carcinoma of thyroid		

A group of heritable syndromes characterized by aberrant growth of benign or malignant tumors in a subset of endocrine tissues have been given the collective term multiple endocrine neoplasia (MEN). The tumors may be functional (ie, capable of elaborating hormonal products that result in specific clinical findings characteristic of the hormone excess state) or nonfunctional. There are three major syndromes: MEN1 is characterized by tumors involving the parathyroid glands, the endocrine pancreas, and the pituitary; MEN 2A includes medullary carcinoma of the thyroid (MCT), pheochromocytoma, and hyperparathyroidism; and MEN 2B, like MEN 2A, includes MCT and pheochromocytoma, but hyperparathyroidism is typically absent.

MULTIPLE ENDOCRINE NEOPLASIA TYPE 1

MEN1, also known as Wermer syndrome, is inherited as an autosomal dominant trait with an estimated prevalence of 2 to 20 per 100,000 in the general population. Approximately 10% of MEN1 mutations arise *de novo*. The term **sporadic MEN1** has been applied to this group. MEN1 has a number of unusual clinical manifestations (Table 22–1) that occur with variable frequency among individuals within affected kindreds. Manifestations of MEN1 have been reported prior to 5 years of age; penetrance is maximal by the fifth decade of life. The diagnosis of MEN1 can

be made based one of three criteria. First, it can be based on the occurrence of 2 or more primary MEN1-associated tumors (discussed later) in a single individual. Second, it can be made in a first degree relative of a clinically diagnosed MEN1 patient if they present with a single MEN1-associated tumor. Third, it can be diagnosed based on identification of a germline MEN1 mutation.

Hyperparathyroidism is the most common feature of MEN1, with an estimated penetrance of 90% or greater over the lifetime of an individual harboring the *MEN1* gene. The diagnosis of hyperparathyroidism is usually made through a combination of clinical and laboratory criteria similar to those used in the identification of sporadic disease (see Chapter 8). It is typically the first clinical manifestation of MEN1, although this varies as a function of the patient population being examined. Hyperparathyroidism in MEN1 is due to hyperplasia of all four parathyroid glands (or more, if supernumerary glands are present). However, involved glands may undergo metachronous enlargement, and selective resection of these glands often results in sustained clinical remissions. MEN1 is a rare cause of hyperparathyroidism, accounting for only 2% to 4% of cases in the general population.

Enteropancreatic tumors in MEN1 (~50% of all MEN1 patients) can be either functional (ie, capable of producing a secreted product with biologic activity) or nonfunctional. Gastrinomas, frequently associated with Zollinger-Ellison syndrome (ZES),

TABLE 22–1 Clinical manifestations of MEN1.

Manifestation	(%)
Hyperparathyroidism	90
Enteropancreatic tumors	30-70
Gastrinoma	40
Insulinoma	10
Non-functioning and PPoma	20-55
Glucagonoma	<1
VIPoma	<1
Pituitary adenomas	30-40
Prolactinoma	20
Somatotropinoma	10
Corticotropinoma	<5
Non-functional	<5
Carcinoid tumors	10-20
Adrenal adenomas	25-40
Subcutaneous lipomas	30
Facial angiofibromas	85
Meningioma	8
Collagenomas	70

Abbreviations: PP, pancreatic polypeptide; VIP, vasoactive intestinal polypeptide.

represent approximately 40% to 60% of the enteropancreatic tumors associated with this syndrome. Of equal importance, roughly 25% of patients with ZES are found in MEN1 kindreds. Insulinomas constitute approximately 20% of the islet cell tumors, and the remainder represent a collection of functional (eg, glucagon- or vasoactive intestinal peptide [VIP]-producing tumors) and nonfunctional tumors. It is noteworthy that the gastrinomas of MEN1 are often small, multicentric, and ectopically located outside the pancreatic bed, most often in the duodenal submucosa. This latter feature can have a major impact on the therapeutic approach to these patients (discussed later). Gastrinomas in MEN1 are frequently malignant, as are their sporadic counterparts; however, for reasons that are only poorly understood, the biologic behavior of these tumors is less aggressive than those found in sporadic disease.

The diagnosis of gastrinomas is based on demonstration of hypergastrinemia in the presence of gastric acid hypersecretion. This latter criterion excludes other more common causes of hypergastrinemia (eg, achlorhydria). When the diagnosis is in question, the secretin stimulation test, which stimulates gastrin secretion from gastrinomas but not from normal tissue, may be employed. *Note*: The availability of secretin for parenteral administration is now quite limited. Histamine receptor blockers and proton pump inhibitors reflexively increase serum gastrin levels and should be discontinued 30 hours and 7 days, respectively, prior to testing to exclude gastrinoma. Others have advocated measurements of multiple gastrointestinal hormones following a standardized mixed meal as the most efficient means of detecting the presence of neuroendocrine tumors in MEN1.

Because of their small size, gastrinomas can be difficult to localize in MEN1. Computed tomography (CT) and magnetic resonance imaging (MRI) may be useful in identifying larger lesions but are typically not helpful in identifying smaller ones. These imaging procedures are useful, however, in demonstrating hepatic metastases when present. The most promising localization techniques studied to date include endoscopic and intraoperative ultrasound, selective arterial secretin injection (followed by hepatic vein sampling for gastrin), and radiolabeled octreotide scanning. Each of these has been employed successfully to identify tumors in studies involving small groups of patients. It is important to recognize, however, that almost half of gastrinomas are not found with preoperative localization studies.

Insulinomas in MEN1 are detected using conventional biochemical testing (see Chapter 18). They can also be difficult to localize given their potential for multicentricity. Endoscopic ultrasound and selective arterial infusions of calcium with hepatic vein sampling (for insulin) have been used successfully to identify lesions in small groups of patients. Metastases are present in up to 50% of patients with MEN1-associated insulinoma, compared with 10% of their sporadic counterparts.

Glucagonomas, VIPomas, and somatostatinomas are diagnosed based on twofold or greater elevations in plasma glucagons, VIP, or somatostatin levels, respectively, in the presence of one or more pancreatic nodules. Nonfunctional tumors are typically diagnosed following imaging studies. Plasma glucagon, chromogranin A or pancreatic peptide levels lack the sensitivity to assist detection of nonfunctional tumors.

Pituitary adenomas occur in approximately 30-40% of patients harboring *MEN-1* gene mutations. The majority of such tumors secrete prolactin, with or without secretion of excess growth hormone, followed by those secreting growth hormone alone, nonfunctional tumors, and those secreting excessive amounts of adrenocorticotropic hormone (ACTH) (Cushing disease). In considering patients with growth hormone excess, it is important to recall that ~75% of reported growth hormone releasing hormone (GRH)-producing pancreatic neuroendocrine tumors have developed in MEN1 patients. A prolactinoma variant of MEN1 has been described (Burin variant). This variant is characterized by an increased frequency of prolactinomas, carcinoids, and hyperparathyroidism and infrequent appearance of gastrinomas in affected kindreds. It does not appear to be associated with a specific mutation of the *MEN1* gene. In one instance it has been linked to a germline mutation in the gene encoding the cyclin-dependent kinase inhibitor p27^{Kip1}, but this has not been confirmed in other patients. Pituitary tumors in MEN1 are rarely malignant. Recent studies suggest that they may be larger and more aggressive than their sporadic counterparts, although this is controversial. There is a high incidence of macroadenoma, they display a propensity for making more than a single pituitary hormone, and there is a suggestion that these tumors do not respond as well to medical therapy as do their sporadic counterparts. Diagnosis and management are similar to that of their sporadic counterparts (see Chapter 4).

Adrenal adenomas, including cortisol-producing adenomas, are seen in MEN1, theoretically making the differential diagnosis of Cushing syndrome in this setting complex (ie, adrenal adenoma

vs basophilic adenoma of the pituitary gland vs ectopic ACTH secretion from a carcinoid tumor, which is also commonly associated with this syndrome). Empirically, most hypercortisolemia in this setting is due to pituitary disease. Pheochromocytomas harboring an MEN1 rather than an MEN2 mutation have been described in seven patients. Thyroid disease has been said to be more common in MEN1; however, with the possible exception of thyroid adenomas, this link remains obscure. Subcutaneous lipomas, skin collagenomas, and multiple facial angiofibromas (particularly on the upper lip and vermillion border of the lip) are seen in 30% to 90% of family members in affected kindreds. Although clinically of little importance, when present, they may prove useful in identifying affected individuals within a kindred and lead to more effective screening (discussed later). Carcinoid tumors are seen with increased frequency in MEN1. They are almost exclusively foregut carcinoids and may be found in the thymus, in the lung (bronchial carcinoids), or in the gastric mucosa. It is likely that chronic hypergastrinemia (eg, that resulting from coexistent ZES and proton pump inhibitor therapy) contributes to the increased incidence of gastric carcinoids in MEN1. For unclear reasons, thymic carcinoids appear more commonly in males, bronchial carcinoids in females. They occasionally secrete hormonal products (eg, ectopic ACTH), are often malignant, and may behave aggressively. Leiomyomas, meningiomas, and ependymomas have been described in MEN1. Disease-specific mortality is due to pancreatic islet carcinoma and malignant thymic carcinoid.

Pathogenesis

MEN1 is inherited as an autosomal dominant trait. Traditional linkage studies localized the defective gene to the long arm of chromosome 11q13. Parallel analyses of DNA from endocrine tumors taken from MEN1 patients demonstrated allelic loss in this area, frequently resulting from large DNA deletions. This raised the possibility that the defective gene was a tumor suppressor gene involved in the control of cellular growth. In this paradigm (Figure 22–1), the inherited defective allele is silent in the presence of a normal, functioning allele on the second chromosome. A subsequent somatic mutation (often a deletion that removes the normal allele) results in a null genotype in which the suppressor gene locus is either absent or defective on both alleles. The high frequency with which such deletions occur is thought to account for the dominant nature of this particular genetic defect. Noteworthy, some of the neuroendocrine tumors seen in MEN1 have failed to demonstrate loss of heterozygosity suggesting that other events might be involved in suppressing the second (ie, normal) allele. Release of the tumor suppressor gene's growth regulatory activity results in a hyperplastic growth response in cells harboring the somatic mutation. This promitogenic state likely provides the substrate for subsequent somatic mutations that result in acquisition of a more aggressive, and at times malignant, phenotype, as occurs in gastrinomas associated with ZES. Recent studies have succeeded in identifying the gene (*MEN1*),
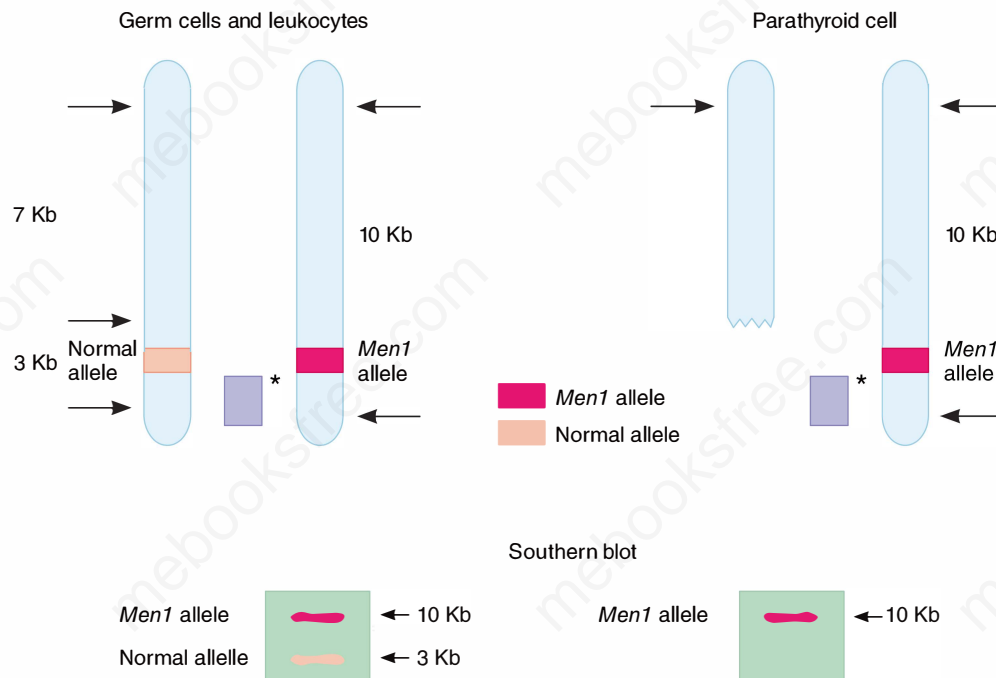
FIGURE 22–1 An example of allelic loss in *Men1* involving large deletion of genetic material. Afflicted patient's germ cells (shown on left) harbor both a normal and defective gene at the *Men1* locus. Each of these is detected using selective restriction enzyme digestion and Southern blot analysis of the genomic DNA. Affected somatic cells (eg, parathyroid chief cells) undergo a second mutation, typically a deletion of the normal allele, resulting in detection of only the mutant allele by Southern analysis. Sporadic disease is thought to follow sequential mutation or deletion of each *Men1* allele in the somatic cell. Solid bar (light purple) with asterisk represents radiolabeled probe used in Southern analysis; arrows indicate points of restriction enzyme digestion.
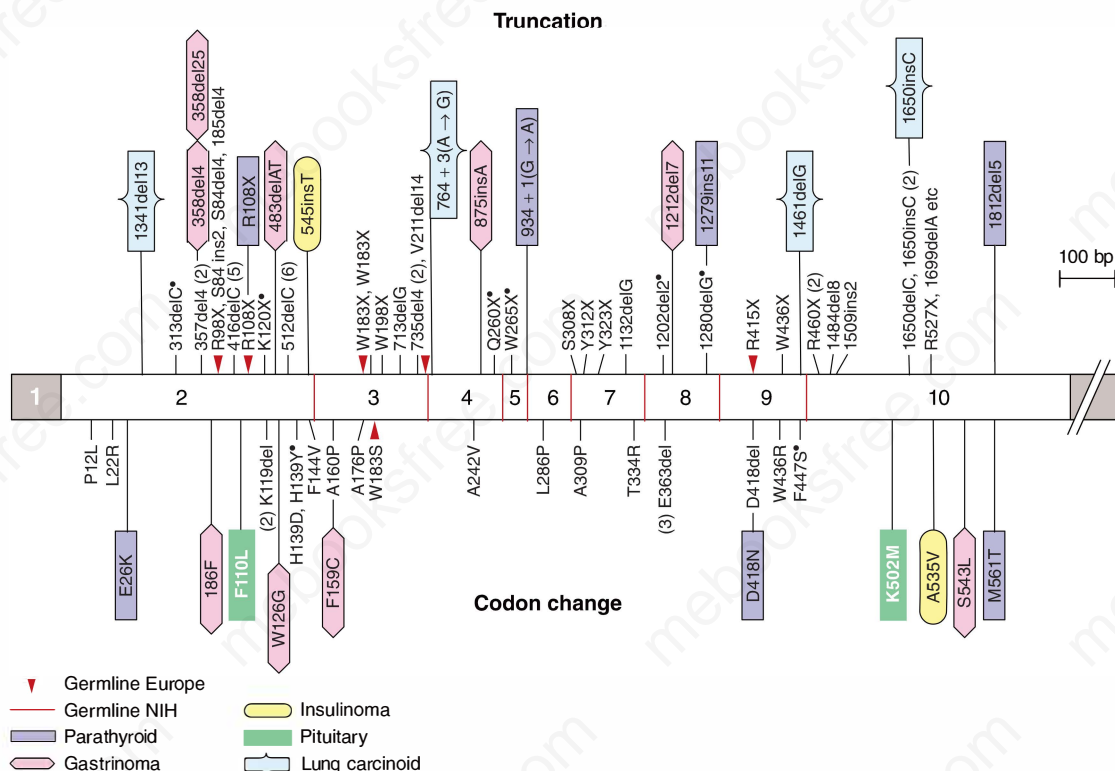
FIGURE 22–2 Some of the more common germline menin mutations from 56 MEN1 kindreds and somatic mutations from 24 endocrine tumors. Mutations above the bar cause protein truncations through stop codons or frameshifts leading to premature stop codons; two cause splice errors. Mutations below the bar cause missense or single-amino-acid codon changes. (Reproduced with permission from Marx S, et al. Multiple endocrine neoplasia type 1: clinical and genetic topics. *Ann Intern Med*. 1998 Sep 15;129(6):484-494.)

encoding the protein menin, which appears to be responsible for MEN1 (Figure 22–2). Mutations have been identified throughout the entire 610-amino-acid length of the *menin* coding sequence and include nonsense mutations, missense mutations, and deletions. Over 1300 independent mutations have been described in MEN1 kindreds to date. Nevertheless, 5% to 10% of presumed *MEN1* mutations are not detectable using available methodology. Menin is a ubiquitously expressed tumor suppressor gene product which has been shown to interact with several components of the trithorax family histone methyltransferase complex, including the mixed lineage leukemia (MLL) proteins. Menin, through this complex, controls expression of the cyclin-dependent kinase inhibitors p27^{Kip1} and p18^{Ink4c}. This complex also methylates the lysine 4 residue of histone 3 (H3K4), a modification which has been linked to transcriptional activation. Loss of function of *menin* (eg, through those mutations associated with MEN1) results in downregulation of p27^{Kip1} and p18^{Ink4c} and deregulated cell growth. Menin also suppresses expression of the developmentally programmed transcription factor, HLXB9, which may also lead to increased growth in menin-deficient cells. Mice heterozygous for deletion of the *menin* gene locus (*Men1-/-*) develop parathyroid dysplasia, adenomas, and carcinomas; pancreatic islet tumors, anterior pituitary tumors, and adrenal cortical tumors, all of which display loss of heterozygosity at the *Men1* locus.

Mutation of a second gene, *CDKN1B*, encoding the cyclin dependent kinase inhibitor p27^{KIP1} (discussed earlier) has been reported in patients who lacked menin mutations but did have the MEN1-associated tumors, particularly parathyroid and pituitary adenomas and, much less commonly, neuroendocrine tumors of the pancreas. This condition has recently been referred to as MEN4 and it accounts for ~1.5%-3% of patients with the MEN1 phenotype. Germline mutations in other cyclin dependent kinase inhibitors, including p15^{INK}, p18^{INK}, and p21 may account for 1%, 0.5%, and 0.5% of MEN1 patients, respectively.

The *MEN1* gene is also the gene most frequently mutated in a variety of sporadic endocrine tumors. Parathyroid adenomas (20%), gastrinomas (40%), insulinomas (15%), VIP-secreting tumors (60%), nonfunctioning pancreatic tumors (15%), glucagonomas (60%), adrenal cortical tumors (<5%), angiofibromas (10%), lipomas (30%), bronchial carcinoids (35%), and pituitary tumors (<5%) often harbor somatic mutations at the *MEN1* locus. It is important to note that screening for germline mutations at the *MEN1* locus fails to detect mutations 5% to 10% of the time despite evidence for loss of heterozygosity at 11q13. Some of these may represent alternate gene targets (discussed MEN4 earlier). In other cases, it may be that the relevant mutations are in regulatory regions surrounding the *MEN1* gene. Haplotype testing or genetic linkage analysis can be useful in these cases to identify relevant *MEN1* carriers. Unlike MEN2 (discussed later), there is no clear

association between specific *MEN1* gene mutations and the nature or extent of endocrine gland involvement, although nonsense mutations at Tyr[312] and Arg[460] have been shown to be more common with the Burin variant.

Treatment

Treatment of MEN1-associated tumors is more complicated than that of its sporadic counterparts. First, MEN1-associated tumors are usually multiple, making surgical cure difficult, if not impossible. Second, occult metastases are more common in MEN1-associated tumors. Third, MEN1-associated tumors are often larger, clinically aggressive and more resistant to therapy than their sporadic counterparts. Therapy of hyperparathyroidism in MEN1 is directed toward surgical extirpation of hyperplastic parathyroid tissue, typically with resection of three and one-half glands. This leaves one-half gland in situ in an attempt to preserve residual parathyroid function and avoid hypoparathyroidism. Alternatively, patients may be subjected to total parathyroidectomy with transplantation of the most normal appearing tissue to the nondominant forearm. There have been no head-to-head comparisons made between these approaches. Total parathyroidectomy (with transplant) is associated with a higher frequency of hypoparathyroidism, whereas subtotal parathyroidectomy has a higher rate of recurrence. Prophylactic near-total thymectomy is often performed at the time of the initial neck exploration to remove potential intrathymic parathyroid glands as well as thymic carcinoid tumors.

Both persistence and recurrence of hyperparathyroidism occur more frequently in MEN1 than in sporadic disease. Persistence of hyperparathyroidism, defined as a failure to normalize serum calcium and parathyroid hormone (PTH) levels following the initial surgery, with removal of 3 to 3.5 parathyroid glands, occurs in 12% of cases of MEN1. Recurrent disease, defined as reappearance of hyperparathyroidism following at least 3 months of normocalcemia, is seen in as many as 44% of these cases 8 to 12 years following surgery. The high frequency of persistent hyperparathyroidism in MEN1 probably reflects the high frequency of supernumerary glands (prevalence as high as 30%) and ectopically located parathyroid tissue in patients carrying the *Men1* gene. The increased frequency of recurrent disease is thought to result from the continued presence of the underlying mitogenic stimulus that drives parathyroid gland growth in this syndrome. Reoperation for recurrent disease in the neck is successful about 90% of the time, whereas autograft removal is successful in about 60% of patients.

Therapy of gastrinomas in MEN1 remains controversial. Suppression of gastric acid production with proton pump inhibitors (eg, omeprazole) remains a mainstay of therapy. Conservative medical management of these tumors has been predicated on their assumed low-grade malignant behavior, the dominance of complications related to gastric acid hypersecretion in contributing to morbidity and mortality, and the failure of most attempts at surgical resection to alter the natural course of the disease. More recently, recognition of the potential for more aggressive behavior in some of these tumors—in a recent study, 14% of these tumors demonstrated aggressive growth—and the fact that many of these

tumors are ectopically located in duodenal submucosa and peripancreatic lymph nodes rather than the pancreatic bed has renewed interest in the possibility of surgical cure. A number of small studies have reported encouraging results when measures to both localize and remove gastrinoma tissue in the pancreas, duodenum, and regional lymph nodes have been used. Although the nature of the underlying genetic lesion and the multicentricity of these tumors may place limits on the prospects for cure of this disease in most patients, the slow growth characteristics of these tumors permit long periods of symptom-free survival following reduction of the tumor burden. For patients with liver or other metastatic disease, symptoms related to hypergastrinemia may be controlled with the proton pump antagonists, as described earlier. More conventional cancer therapy (eg, systemic chemotherapy, radiation therapy, or selective chemoembolization of hepatic metastases) is palliative and reserved for advanced stages of the disease. Islet tumors more than 3 cm carry an increased risk of malignancy and are usually resected regardless of functional status.

It is important to remember that calcium stimulates gastric acid secretion. This may occur through gastrin-dependent and gastrin-independent pathways. In MEN1 patients with both hyperparathyroidism and ZES, correction of the hyperparathyroidism and attendant hypercalcemia frequently results in a reduction in both basal and maximal acid output and a decline in serum gastrin levels. Secretin stimulation tests often normalize following parathyroidectomy. More importantly, there is a reduction in the dose of medication (eg, proton pump inhibitor) required to control symptoms of ZES following parathyroid surgery in approximately 60% of patients.

Unlike gastrinomas, insulinomas are rarely localized outside the pancreatic bed. Therefore, a more aggressive approach can be directed toward the pancreas in planning surgical resection. Enucleation of the identifiable lesions in the pancreatic head and blind resection of the pancreatic body and tail are often successful in correcting hyperinsulinemia. Resection of glucagonomas and VIPomas, regardless of size, has also been recommended for symptom control. Nonsurgical candidates (eg, those with serious coexisting disease or those in whom candidate tumors cannot be identified) can be managed with conventional medical therapy (eg, diazoxide or verapamil) (see Chapter 18).

Nonfunctional tumors may be the most common pancreatoduodenal neuroendocrine tumors in MEN1. They are managed somewhat differently than the functional tumors. Surgical efforts to minimize spread of metastases are weighed against the need to avoid incurring surgical complications, including mortality, with little clinical benefit. In general, surgery is not recommended for tumors less than 2 cm in diameter. Life expectancy of MEN1 patients with nonfunctional tumors (~43 years) is similar to that of those with gastrinomas and considerably shorter than those with other types of functional tumors (~61 years).

Screening

MEN1 gene testing should be offered to all index cases and their first-degree relatives. Guidelines suggest that this should be done

at the earliest opportunity since clinical manifestations can appear before the age of 5. MEN1 accounts for less than 1% of all pituitary tumors, 2% to 4% of cases of primary hyperparathyroidism, 20% of all thymic neuroendocrine tumors, and about 25% of all gastrinomas. Thus, although routine screening for *Men1* gene mutations is not indicated for sporadic cases of hyperparathyroidism or patients with pituitary tumors, screening of all cases of ZES and thymic neuroendocrine tumors is likely to be cost-effective in identifying carriers. The frequency of germline mutations in the *Men1* gene in patients with tumors thought to be sporadic based on family analysis is about 5% for gastrinomas and 1% to 2% for other manifestations (eg, hyperparathyroidism, prolactinomas). Excluding those patients with ZES or thymic neuroendocrine tumors, screening for MEN1 outside affected MEN1 kindreds should be limited to those individuals for whom the index of suspicion for the syndrome is high (eg, familial history of endocrine tumors or hypersecretory states; history of multiple endocrine tumors, including pancreatic neuroendocrine tumors, or parathyroid hyperplasia at age <30 in the index case). Identification of the carrier state should be performed with the intent of acquiring information that allows the clinician to focus screening on the relevant patient population (eg, members of affected kindred who do not share the *Men1* gene mutation need not be subjected to follow-up screening). Unlike the situation in MEN2 (discussed later), carrier analysis should not be used to support a major therapeutic intervention. Such interventions (eg, pancreatic exploration) may be associated with significant morbidity, and there is no evidence that they prolong patient survival. Genetic screening is available commercially at a number of locations (see www.genetests.org).

Although most clinicians agree that screening for MEN1-associated neoplasia in high-risk groups (eg patients with MEN1 and relatives harboring a mutant MEN1 allele) is worthwhile, details of the individual screening protocols vary considerably—from more focused, cost-effective approaches to broad-based screening designed to detect occult disease. One example is presented in Figure 22–3. A complete examination and full family history should be obtained on all patients. In individuals within affected kindreds, ionized (or albumin-corrected) calcium and PTH levels should be checked at yearly intervals beginning around age 8. Fasting serum gastrin levels (with documentation of gastric hyperacidity) should also be determined annually beginning at age 20—or more frequently if ZES is a prominent component of the phenotype in the affected kindred. In kindreds in whom the disease is particularly aggressive, the secretin stimulation test (if available) can lend additional diagnostic sensitivity. Insulinoma screening with determination of fasting glucose, insulin, and proinsulin levels may prove useful, particularly if symptoms of hypoglycemia are present. This screening should begin around age 5. Pancreatic polypeptide measurements may identify an unsuspected pancreatic neuroendocrine tumor. Levels of chromogranin A, a secretory product of parathyroid glands and other neuroendocrine tissues, can also be helpful in assessing hypersecretory states, although elevations are not specific for MEN1-associated endocrinopathies. Routine screening for other functional islet cell tumors (eg, glucagonomas and VIPomas) is justified if clinical findings (eg, watery diarrhea, hypokalemia, necrolytic migratory erythema) are present. As noted earlier, pancreatic neuroendocrine lesions are typically multifocal and may be very small, which make detection difficult even with sophisticated imaging studies. The same holds true for the pituitary lesions. In the absence of obvious clinical findings (eg, evidence of a hyper- or hyposecretory state or symptoms referable to a mass lesion in the sella turcica), routine screening should be confined to periodic measurements of serum prolactin and perhaps insulin-like growth factor-I. The former has been found to be useful in identifying pituitary disease in females harboring the *Men1* gene defect. Imaging studies (eg, CT or MRI of the pituitary, anterior mediastinum and abdomen) should be performed at presentation and repeated at 1-3-year intervals, based on perceived clinical risk for developing neoplastic lesions. A thorough family history that excludes pituitary disease in the kindred may mitigate the need for pituitary imaging studies over the long term.

Penetrance of MEN1 is greater than 95% by age 45. Screening should be continued at periodic intervals at least to age 45. If there is no evidence of typical endocrine organ involvement by that age, screening frequency might be reduced. It is important to note, however, that the risk is not reduced to zero at age 45. A minority of patients presents with their first manifestation of the syndrome well after age 45. Surgical resection of diseased tissue (eg, parathyroidectomy) should be followed with continued screening looking both for recurrent disease and involvement of other organ systems.

MULTIPLE ENDOCRINE NEOPLASIA TYPE 2

MEN2 is an autosomal dominant disorder with an estimated prevalence of 1 to 10 per 100,000 in the general population. It can be subdivided into two independent syndromes: MEN 2A (Sipple syndrome) and MEN 2B. Manifestations of MEN 2A include MCT, pheochromocytoma, and hyperparathyroidism. MEN 2B includes MCT, pheochromocytoma, and a number of somatic manifestations (Table 22–2; Figure 22–4), but hyperparathyroidism is rare.

Medullary carcinoma of the thyroid is the most common manifestation of MEN2 and often represents the first clinical presentation in individuals with multiorgan involvement. It also dominates the clinical course of patients affected with the disease. From 80% to 100% of individuals at risk develop MCT at some point during their lifetime. The classic thyroid lesion of MEN2 is hyperplasia of the calcitonin-producing parafollicular cells, which typically serves as the precursor of MCT. These tumors in MEN2 patients tend to be multicentric and concentrated in the upper third of the thyroid gland, reflecting the normal distribution of parafollicular cells.

As much as one-fourth of all MCT is genetic in origin. Roughly 80% of the heritable fraction is attributable to *MEN 2A*; 15% occurs as an isolated entity (isolated familial MCT), and 5% is found in *MEN 2B* kindreds. Isolated familial MCT is defined as MCT presenting as the only clinical manifestation in more than
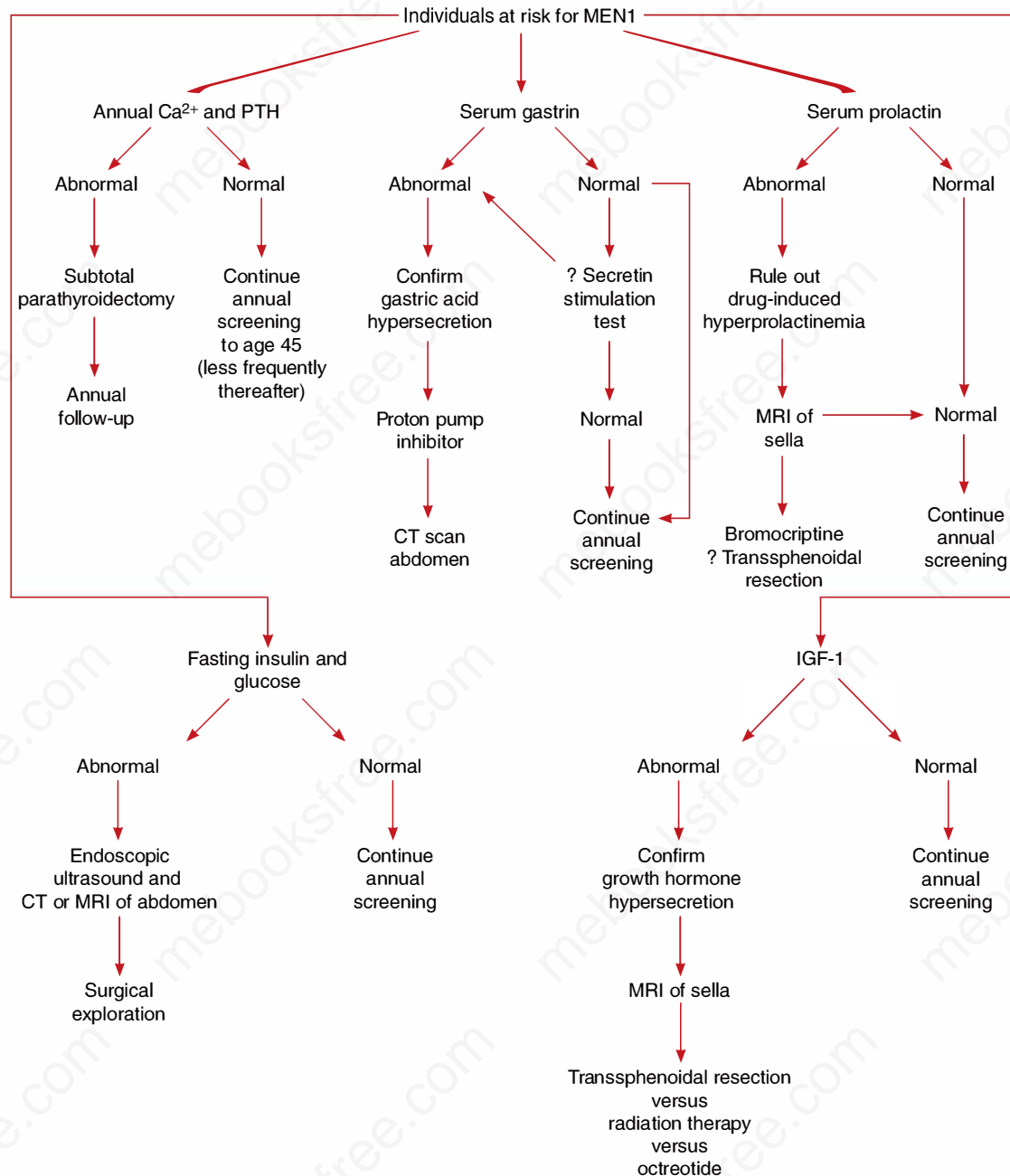
FIGURE 22–3 Screening for MEN1-associated neoplasia. Testing is targeted to individuals who either harbor the *MEN1* gene or have a high probability of harboring the *MEN1* gene (eg, first-degree relatives of affected kindred members). If genetic testing is available and the specific mutation has been confirmed, screening for endocrine neoplasia would be limited to those individuals harboring the defective gene. Screening for other pancreatic (eg, insulinomas) or pituitary tumors (eg, somatotroph adenomas) is based on the prevalence of the specific lesion in the affected kindred or the presence of signs or symptoms suggesting a particular lesion (IGF, insulin-like growth factor).

10 carriers in a kindred and multiple carriers or affected members over the age of 50 with adequate history and laboratory support to rule out the presence of pheochromocytoma or hyperparathyroidism. The disease tends to behave more aggressively in MEN 2B than with either MEN 2A or familial MCT, with earlier presentation (often before age 5) and more rapid progression.

Biochemical diagnosis depends heavily on the calcitonin-producing properties of the hyperplastic parafollicular cells or MCT. These lesions respond to pentagastrin or calcium infusions with significant increments in plasma calcitonin levels (see Chapter 8). Occasionally, immunohistochemical staining of poorly differentiated thyroid tumors for calcitonin reveals the identity of the malignancy. The presence of extracellular amyloid is also one of the identifying features of these tumors. This material reacts with anticalcitonin antisera, suggesting that it includes aggregated hormone released from neighboring tumor cells. MCT spreads

TABLE 22–2 Clinical manifestations of MEN2.

Manifestation	(%)
MEN 2A	
Medullary carcinoma of thyroid	80-100
Pheochromocytoma	40
Hyperparathyroidism	25
MEN 2B	
Medullary carcinoma of thyroid	100
Pheochromocytoma	50
Marfanoid habitus	75
Mucosal neuromas	100
Ganglioneuromatosis of bowel	>40

initially within the thyroid bed and to regional lymph nodes. Distant metastases to liver, lung, and bone occur late in the course of the disease.

Pheochromocytomas develop in approximately 50% of individuals harboring the *MEN2* gene. Pheochromocytoma is diagnosed after MCT in 50% of cases, concurrently in 40% of cases and before MTC in 10% of cases. They are usually located in the adrenal bed, often are bilateral, and are rarely malignant. Diagnosis is based on standard clinical criteria (eg, hypertension, presence of headaches, palpitations, diaphoresis), elevations in plasma or urine catecholamines or catecholamine metabolites (eg, urinary or plasma metanephrine or normetanephrine), and demonstration of an adrenal mass on conventional abdominal CT scanning or MR imaging. As noted earlier for MCT, pheochromocytomas in MEN2 are preceded by a hyperplastic phase (adrenal medullary

FIGURE 22–4 Patient with MEN 2B syndrome. Note the multiple neuromas on the lips and tongue and the marfanoid facies.

hyperplasia), although, unlike parafollicular cell hyperplasia, the adrenal precursor lesion can be difficult to detect with conventional biochemical testing. As in MEN1, hyperparathyroidism in MEN2 is due to hyperplasia of the parathyroid glands. It is seen in about 25% of patients harboring the *MEN 2A* gene and is reported, but rarely seen, as part of MEN 2B. The disease is usually less aggressive than its counterpart in MEN1 and approximates more closely the behavior of sporadic disease. It responds well to surgical management.

A number of other phenotypic features are associated with the MEN2 syndromes. Cutaneous lichen amyloidosis is a pruritic erythematous skin lesion that is seen coincident with or often preceding the development of MCT in MEN2. Amyloid in these lesions is composed of keratin rather than calcitonin, as seen in MCT. Cutaneous lichen amyloidosis is thought to result from an abnormality in cutaneous innervation that leads to hyperesthesia and pruritis. Repeated scratching of the pruritic area results in epidermal thickening and increased pigmentation. It has been noted more frequently in association with specific mutations of the *MEN2* gene (specifically Cys[634] and less commonly Val[804]). It can be present before the onset of clinically evident MCT. In a second variant, MEN 2A or familial MCT is associated with Hirschsprung disease (congenital megacolon). This is most frequently found with *RET* gene mutations involving Cys[609] (15%), Cys[611] (5%), Cys[618] (30%), and Cys[620] (50%) (see later). The intestinal ganglioneuromatosis, pes cavus, pectus excavatum, mucosal neuromas, marfanoid habitus, and medullated corneal nerves seen in MEN 2B appear to be related to the underlying genetic defect. The intestinal lesions can disrupt gut motility, resulting in periods of severe constipation or diarrhea.

Pathogenesis

The pathogenesis of the MEN2 syndromes has been worked out in elegant detail. Traditional genetic linkage studies localized the defective genes in *MEN 2A*, *MEN 2B*, and familial *MCT* to the pericentromeric region of chromosome 10. Subsequent refinement in these analyses indicated that the defective gene was either closely linked to or identical with the *ret* proto-oncogene. RET is a single transmembrane domain, tyrosine kinase-linked protein that forms part of the receptor for the glial cell line–derived neurotrophic factor (GDNF) (Figure 22–5). This receptor, GDNFR (glial cell line–derived neurotrophic factor receptor), is a glycosyl phosphatidyl inositol-linked cell-surface protein. Additional ligands, neurturin, persephin, and artemin have been shown to associate with and activate RET through their cognate receptors GDNFR$_{\alpha1-4}$. As depicted schematically in Figure 22–6, its most striking structural feature is a series of cysteine residues clustered just outside the membrane-spanning segment. These cysteine residues are thought to exert a tonic inhibitory control on RET activity in the normal cell. RET also has a cadherin-like domain in that portion of the molecule projecting into the extracellular space and a tyrosine kinase-like domain in the intracellular portion of the molecule. RET is expressed endogenously in a variety of cells of neural crest origin,
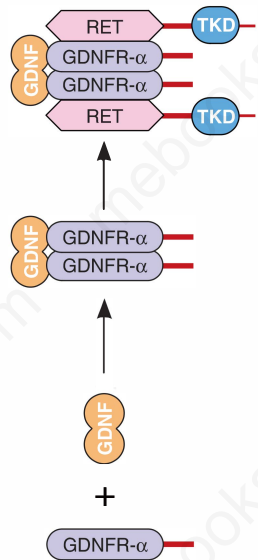
FIGURE 22–5 Structure of GDNFR-α/RET receptor complex. GDNF identifies the ligand, glial cell line–derived neurotrophic factor, and GDNFR-α, the GDNF receptor (TKD, tyrosine kinase domain).
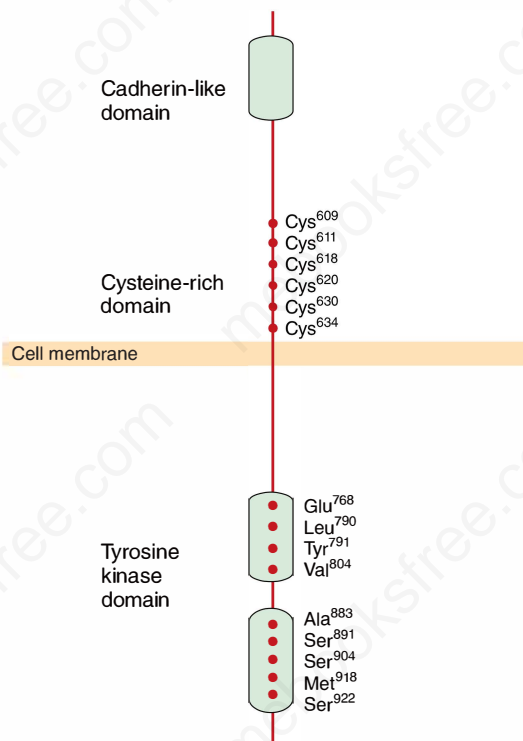
FIGURE 22–6 Structural schematic of wild-type RET, including amino acids that have been shown to be mutated in different disease states. MEN 2A is associated with mutations of Cys^{609}, Cys^{611}, Cys^{618}, Cys^{620}, Cys^{630}, Cys^{634}, and, rarely, Leu^{790} and Tyr^{791}. Familial MCT of the thyroid is frequently associated with mutations at the same Cys residues with the exception of Cys^{634}, as well as mutations at Glu^{768}, Leu^{790}, Tyr^{791}, Val^{804}, and Ser^{891}. MEN 2B mutations typically involve Met^{918} or, more rarely, Ala^{883}, Ser^{922}, or Val^{804}/Ser^{904}. Hirschsprung disease is associated with multiple mutations or deletions extending over the full length of the RET molecule.

and it appears to play an important role in development. Knockout of the *RET* gene locus in mice results in the absence of myenteric ganglia in the submucosa of the small and large intestine and a variety of genitourinary anomalies, implying an important role in renal development. Approximately 95% of MEN 2A and 2B patients and 88% of familial MCT have identifiable *RET* mutations. From 1% to 7% of sporadic MCT have somatic mutations in the *RET* gene.

Characterization of the *RET* gene in patients with MEN 2A demonstrated a number of mutations in affected kindred members that were not present in their normal counterparts. The mutations were clustered in the cysteines located in RET's extracellular, juxtamembrane domain (see Figure 22–6). Structurally, these cysteines are encoded by nucleotides in exons 10 and 11 of the *RET* gene. A number of these mutations were simple missense mutations, whereas others involved deletion or insertion of small segments of DNA, but in each case one of the aforementioned cysteines proved to be involved. Selective mutation of one of these six cysteines has now been shown to account for more than 97% of all *RET* mutations associated with MEN 2A. The most frequently mutated residue is Cys^{634}. This amino acid is mutated to arginine (Arg), phenylalanine (Phe), serine (Ser), glycine (Gly), tyrosine (Tyr), or tryptophan (Trp) in approximately 84% of affected MEN 2A kindreds. It should be noted that the Cys-to-Arg mutation at position 634 is also the most common mutation at this position, accounting for about 64% of all codon changes at this location. Interestingly, a rare but unique mutation involving a four-amino-acid insertion between Cys^{634} and Arg^{635} has been described that results in MCT and a high incidence of hyperparathyroidism but not pheochromocytoma.

Ironically, patients with familial MCT display many of the same cysteine mutations identified in MEN 2A. This implies the existence of independent modulatory genomic or epigenetic factors that restrict the effects of the *RET* mutation to the parafollicular cells of the thyroid in familial MCT. The exception is the mutation of Cys^{634} to Arg^{634}, which is almost uniformly associated with MEN 2A. Several additional mutations have been identified that may be more specific for familial MCT (see Figure 22–6). The observation that mutations of Cys^{634} have higher transforming potential than mutations at these other residues has led to the suggestion that MEN 2A represents the more severe phenotype (vs familial MCT) along a spectrum of disease resulting from RET activation. In fact, the noncysteine mutations, which are more common in familial MCT (as high as 60% of familial MCT in one series), are associated with delayed onset of parafollicular C cell disease, but other clinical features (tumor size, bilaterality, the presence of nodal metastases) are not different in patients harboring cysteine versus noncysteine mutations.

Interestingly, patients with MEN 2B do not harbor mutations in the cysteine residues affected in MEN 2A and familial MCT. Instead, the majority possess a single-point mutation involving conversion of Met^{918} to Thr^{918}. This mutation is found in more than 95% of cases of MEN 2B. A minority of patients have been shown to harbor an independent mutation of Ala^{883} to Phe^{883},

Ser^{922} to Tyr^{922}, or a mono-allelic combination of Val^{804} to Met^{804} plus Ser^{904} to Cys^{904} or Asp^{805} to Lys^{805}. MEN 2B arises from spontaneous *de novo* mutations in as many as 50% of affected individuals. For unknown reasons, these mutations are found almost exclusively on the paternal allele.

A number of the germline mutations identified in MEN 2A and familial MCT and the Met^{918} mutation in MEN 2B have been demonstrated as somatic mutations in sporadic MCT (~30%-40%) and pheochromocytoma (<10%). Sporadic parathyroid disease due to these mutations is not known to occur. The presence of the Met^{918} mutation, in particular, is associated with a less favorable clinical outcome in sporadic MCT.

Independent studies have shown a close linkage between the *RET* gene locus and Hirschsprung disease, a disorder characterized by failure of myenteric ganglia to develop normally in the hindgut of affected individuals. This leads to impaired gut motility and, in severe cases, megacolon (a phenotype similar to that reported for the *RET* knockout mice). Examination of *RET* coding sequence in Hirschsprung kindreds revealed a variety of mutations in both the intracellular and extracellular domains, some of which (eg, deletions) would be predicted to eliminate normal expression of the *RET* gene. This, together with findings from the *RET*, *GDNF*, and *GDNFRa-1* knockout mice, suggests that Hirschsprung disease represents the null phenotype for the *RET* locus. Interestingly, several patients have been described who possess features of both MEN2 and Hirschsprung disease. The *RET* mutations in these cases have involved conversion of Cys^{609}, Cys^{611}, Cys^{618}, or Cys^{620} to Arg. These mutations, while promoting dimerization and increasing tyrosine kinase activity in the RET protein (discussed later), also appear to have difficulty trafficking to and accumulating in the plasma membrane. It is likely that predominance of one or the other of these mechanisms in different cell types could result in a phenotype characterized by both activation (eg, MEN2) and suppression due to haploinsufficiency (eg, Hirschsprung disease) of RET activity in the same individual.

By inference, the defect in RET function in MEN2 or familial MCT arises from increased or altered activity of the RET tyrosine kinase. In the case of RET$_{MEN 2A}$, the increase in activity appears to arise from interference with the tonic inhibition of RET tyrosine kinase activity by the clustered cysteine residues in the extracellular domain. This leads to increased dimer formation, autophosphorylation, and tyrosine kinase activity in the mutant RET molecules. In the case of RET$_{MEN 2B}$, there appears to be a constituitive increase in tyrosine kinase activity as well as a change in substrate specificity of the tyrosine kinase that contributes to the phenotype. The activity of RET$_{MEN 2B}$—rather than being restricted to conventional RET substrates (RET substrates are similar to those recognized by the epidermal growth factor receptor)—is capable of phosphorylating substrates normally recognized by members of the Src and Abl families of cytoplasmic tyrosine kinases, signaling pathways which are closely identified with the regulation of cell growth. Thus, it appears that RET$_{MEN 2B}$ has acquired the capacity for activation of a potent mitogenic pathway in the expressing endocrine cells merely by altering its selection of substrates for phosphorylation. RET$_{MEN 2B}$ also potentiates phosphorylation of Tyr^{1062} more effectively than RET$_{MEN 2A}$. This tyrosine serves as a docking site for multiple effector proteins, including Shc and phosphatidyl inositol 3′-kinase, implying that RET$_{MEN 2B}$ may be more effective in triggering downstream signaling pathways.

Treatment

Treatment of heritable MCT should include total thyroidectomy with at least central lymph node dissection. Given the multicentric nature of the disease, subtotal thyroidectomy predictably results in recurrent disease. Basal or stimulated calcitonin levels are used in the postoperative setting to evaluate the presence of residual disease. Postoperative measurement of serum thyroglobulin levels provide an estimate of residual thyroid tissue and, inferentially, tissue at risk for developing MCT. The precise timing of surgery in patients with subclinical disease (ie, positive by genetic testing but without clinical or laboratory abnormalities) is controversial (discussed later), but most clinicians would agree that in kindreds with MEN 2B or clinically aggressive MCT, patients should undergo surgery as soon as the genetic defect is demonstrated. Typically, this is before age 6 months in MEN 2B (eg, mutation at Ala^{883} or Met^{918}) and before age 5 years in MEN 2A (eg, mutation at Cys^{611}, Cys^{618}, Cys^{620}, or Cys^{634}), although some would argue that Cys^{634} should be operated in the first year of life. Foci of microscopic MCT are common, and metastatic disease has been described in the first year of life in patients with MEN 2B. The presence of a thyroid nodule or lymph node metastases at the time of diagnosis is strongly associated with persistent or recurrent disease and low rate of cure. Mutations associated with familial MCT (eg, Cys^{609}, Glu^{768}, Leu^{790}, Tyr^{791}, Val^{804}, or Ser^{891}) tend to be associated with more indolent disease. Some have suggested that surgery in these patients can be deferred until plasma calcitonin levels (basal or stimulated) are elevated. Patients should always be screened for the presence of pheochromocytoma before undergoing neck exploration. Surgery for metastatic disease is palliative and targeted at reducing tumor burden rather than cure. Localization techniques (eg, MRI or selective venous sampling for calcitonin) can be helpful in identifying foci of malignant tissue. Recent randomized, double blind, phase 3 clinical trials involving the tyrosine kinase inhibitors vandetanib and cabozantinib have shown significant improvement in progression-free survival in patients treated with these agents. Each has now been approved for treatment of patients with advanced MCT. These agents (vandetanib and cabozantinib) provide objective response rates of 45% and 28%, respectively, when compared with placebo. Radiation and chemotherapy are of limited utility in MCT and are largely confined to later stages of the disease.

Treatment of pheochromocytomas in MEN2 is similar to that for sporadic pheochromocytomas (see Chapter 11). Alpha- and (occasionally) beta-adrenergic blockade is used to control blood pressure and associated hyperadrenergic symptoms and to restore normal intravascular volume in preparation for surgical resection of the tumor. Given the propensity for bilaterality of pheochromocytomas in this disorder, some have favored bilateral adrenalectomy at the time of initial surgery. However, because the incidence

of bilaterality is well under 100% and because these tumors are rarely malignant, the most prudent strategy in the face of unilateral adrenal enlargement would appear to be unilateral adrenalectomy at the initial surgery with careful attention at follow-up, looking for the presence of disease in the contralateral adrenal gland. This vigilant approach has the advantage of minimizing morbidity from recurrent pheochromocytoma while sparing the patient the risks associated with lifelong adrenal insufficiency.

Screening

Genetic screening for MEN 2A, MEN 2B, or familial MCT is routinely carried out using polymerase chain reaction (PCR)-based tests designed to identify specific mutations in the *RET* coding sequence (Figure 22–7; see www.genetests.org for laboratories performing these tests). Known *RET* mutations account for more than 95% of all instances of MEN2, and selected mutations (eg, Cys^{634} to Arg^{634} in MEN 2A) account for a disproportionate number of affected individuals. Individuals lacking any of the known *RET* mutations can be tested using conventional haplotype analysis if informative genetic markers and affected family members are available. Biochemical testing using basal or stimulated plasma calcitonin levels has been largely supplanted by genetic screens. The biochemical tests remain useful, however, in identifying residual disease after thyroidectomy.

In view of the fact that a high proportion of cases of MCT are familial to begin with and as much as 3% to 7% of patients with apparently sporadic MCT harbor germline *RET* mutations, genetic testing for *RET* germline mutations is probably indicated for all patients presenting with MCT. Controversy persists, however, in terms of what should be done for patients once the mutation has been identified. Some investigators citing incomplete clinical penetrance (according to published data, 40% of gene carriers do not present symptomatically prior to age 70) have argued that employing solely genetic criteria in making the decision for operative intervention subjects a small minority of patients to premature thyroidectomy. They argue that genetic testing should be used to identify those patients who require close clinical and biochemical surveillance to assist with the timing of surgery. Ideally, such biochemical testing (eg, pentagastrin stimulation)
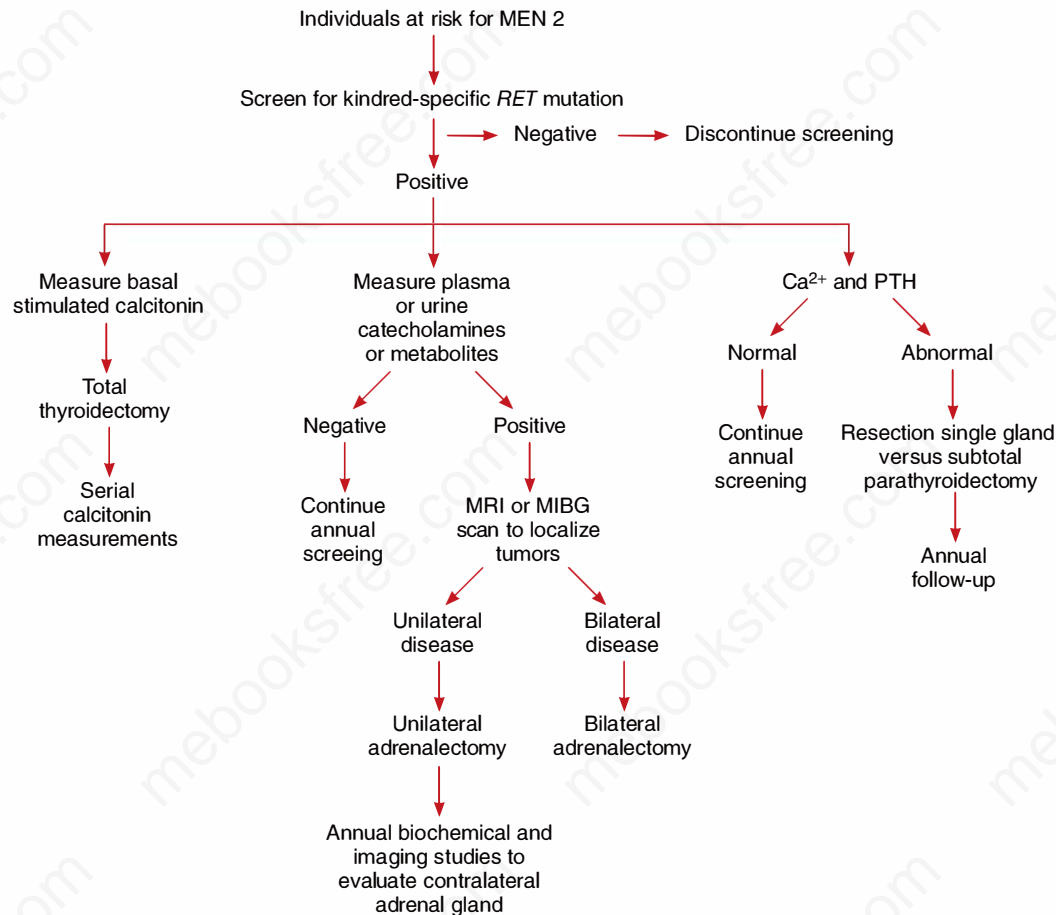
FIGURE 22–7 Screening for neoplasia in MEN 2. Genetic screening has largely supplanted biochemical testing in identifying individuals at risk. Details of decisions regarding treatment are discussed in the text. Preoperative plasma calcitonin levels are potentially useful for follow-up but should not supplant genetic testing in assessing the need for surgery (MIBG, metaiodobenzylguanidine; *RET*, rearranged during transfection proto-oncogene).

should be performed on an annual basis. Exceptions to this general approach might include patients with MEN 2B or a particularly aggressive form of familial MCT where the potential for significant morbidity and mortality would justify operation in any patient harboring the genetic defect regardless of the physical or biochemical manifestations of the disease. Genetic considerations for patients presenting with seemingly sporadic pheochromocytoma are discussed in Chapter 11. *RET* gene mutations in sporadic hyperparathyroidism are very rare, arguing against routine screening for mutations in this group.

Most investigators in the field believe that the true penetrance of MCT—combined clinical and preclinical disease—in MEN 2A is close to 100%. This, when coupled with the high degree of sensitivity and specificity of the PCR-based genetic screens, difficulties encountered in obtaining adequate long-term patient follow-up and biochemical screening, and the potential for false positive and false negative biochemical screening tests (ie, basal or stimulated calcitonin levels)—even within MEN2 kindreds—has led to the recommendation that total thyroidectomy should be performed in most individuals harboring an MEN2–associated *RET* mutation. This argument has now been supported by several clinical studies in which parafollicular cell hyperplasia as well as early MCT have been identified in operative specimens taken from genetically affected individuals despite normal pentagastrin stimulation tests. False-positive biochemical tests are also a concern. There are several reports in the literature of patients in affected kindreds who have undergone total thyroidectomy following positive pentagastrin stimulation tests but did not, in fact, harbor the *MEN2* gene mutation. Histologic examination of

excised tissues revealed parafollicular cell hyperplasia, presumably unrelated to MEN2, but no MCT. Collectively, these findings point out the relative deficiencies of biochemical versus genetic testing and offer a compelling argument for early operation as a means of reliably eradicating the disease.

Several groups have published recommendations regarding the appropriate age to initiate screening and the timing of surgery for MCT in individuals with confirmed *RET* mutations. These are generally segregated based on the nature of the specific *RET* mutations and the relationship of these mutations to the development of aggressive MCT. One set of recommendations is presented in Table 22–3. As noted, prophylactic surgery is recommended for all MEN 2B patients (eg, M918T or A883F). Patients at the next level (ie, C634R/F/S/W/Y) should be screened in the first years of life and operated before age 5. Patients at the third level (basically MEN 2A patients other than the C634 mutations—ie, C609, C611, C620, C630) should be screened early and operated prior to age 5, but more flexibility is available if selected criteria are met (eg, normal calcitonin levels, normal ultrasound, less aggressive MCT in other family members and family preference). The final group, which is associated with a less aggressive clinical course, includes all those mutations associated with familial, isolated MCT. Screening for MCT in this group, as well as consideration for thyroidectomy, can often be deferred beyond age 5 if the criteria above are satisfied. Routine screening for pheochromocytoma and hyperparathyroidism in individuals at risk can typically be initiated later, but all patients undergoing thyroidectomy should have pheochromocytoma ruled out first. Screening should continue at 1-3 year intervals at least until age 50.

TABLE 22–3 Recommendations for timing of prophylactic thyroidectomy and screening for pheochromocytoma and hyperparathyroidism in MEN 2 based on genotype.

ATA Risk Level	RET Mutation(s)	Recommended Age	Pheo Screening	hPTH Screening Thyroidectomy
A (low)	Glu 786	May delay beyond age 5 if calcitonin levels nl, U/S nl and family history of indolent disease	Age 20	Age 20
	Leu 790			
	Val 804			
	Arg 833			
B	Cys 609	Before age 5 (may consider delay based on same criteria used for Risk Level A)	Age 8 for Cys 630	Age 8 for Cys 630
	Cys 611		Age 20 for all others	Age 20 for all others
	Cys 618			
	Cys 620			
	Cys 630			
C	Cys 634	Before age 5	Age 8	Age 8
D (high)	A 833	Early year 1	Age 8	N/A
	M 918			

Mutations described for Risk Level A are representative but do not include all mutations described for this group. Data from Moline J, Eng C. Multiple endocrine neoplasia type 2: an overview. *Genet Med*. 2011 Sep;13(9):755-764.

OTHER DISORDERS CHARACTERIZED BY MULTIPLE ENDOCRINE ORGAN INVOLVEMENT

Carney Complex

Carney complex is an autosomal dominant disorder characterized by cardiac, endocrine, cutaneous, and neural tumors. Myxomas of the heart, breast, and skin are seen frequently in this disorder, as is spotty pigmentation of the skin (lentiginosis) and ear canal trichofolliculo-epitheliomas. Additional rare manifestations include psammomatous melanotic schwannoma, breast ductal adenoma, and osteochondromyxoma. Endocrine tumors include primary pigmented micronodular adrenocortical hyperplasia (an ACTH-independent form of Cushing syndrome), follicular thyroid carcinomas, adrenocortical carcinoma, somatotroph (growth hormone–secreting) adenomas of the pituitary gland, large-cell calcifying Sertoli cell tumors of the testes, and ovarian cysts. Pancreatic neoplasms occur in ~2.5% of patients but may be aggressive. They are the major cause of cancer death in Carney Complex patients. In approximately two-thirds of the known Carney complex kindreds, the genetic lesion maps to 17q22-24, a locus that harbors the type 1A regulatory subunit of protein kinase A. This gene functions as a classic tumor suppressor, and loss of heterozygosity at this locus is associated with the Carney phenotype. A second locus, presumably accounting for some fraction of the remaining half of Carney complex kindreds, is located at 2p16. The nature of the genetic lesion at this locus remains unknown at present; however, the phenotype is indistinguishable from that found with the 17q22-24 mutations alluded to above.

McCune-Albright Syndrome

The central findings in McCune-Albright syndrome include bony fibrous dysplasia (either monostotic or polyostotic, café au lait spots, and gonadotropin-independent precocious puberty). However, affected individuals also develop tumors or nodular hyperplasia of a number of endocrine glands with associated hypersecretory syndromes. One may see involvement of the pituitary somatotrophs (acromegaly), thyroid gland (hyperthyroidism), or the adrenal cortex (hypercortisolemia). Patients with McCune-Albright syndrome frequently demonstrate hypophosphatemic rickets. This is thought to result from secretion of fibroblast growth factor-23 that is produced in the mesenchymal cells associated with the dysplastic bone lesions. They also have nonendocrine organ involvement that leads to cardiomyopathy with sudden death and liver function abnormalities. The molecular lesion in McCune-Albright syndrome involves a somatic activating mutation of *GNAS*, encoding the alpha subunit of the Gs protein, at the Arg^{201} position, converting the Arg to either Cys or His (see Chapter 1), that leads to ligand-independent activation of adenylyl cyclase and increased intracellular concentrations of cyclic AMP. This somatic mutation is thought to occur early in development, leading to genetic mosaicism for the mutant allele. The phenotype of an individual patient is determined by the tissue distribution of the mutation-bearing cells and the percentage of cells within a given tissue that harbor the mutation. Interestingly, *GNAS* mutations have been described in a subset of patients harboring intramuscular myxomas (Mazabraud syndrome) with or without associated fibrous dysplasia, a feature that is reminiscent of the Carney complex (discussed earlier), another genetic disease which seems to target the cyclic AMP-signaling pathway.

Neurofibromatosis Type 1

Neurofibromatosis type 1 (von Recklinghausen disease) is an autosomal dominant genetic disorder characterized by a variety of skin manifestations, including café-au-lait spots, subcutaneous neurofibromas, and axillary and inguinal freckles as well as neural gliomas (eg, optic nerve) and hamartomas of the iris (Lisch nodules). In addition, patients may have a number of endocrine neoplasias, including pheochromocytoma, hyperparathyroidism, MCT, and somatostatin-producing carcinoid tumors of the duodenal wall. The genetic lesion in neurofibromatosis type 1 is located at 17q11.2, a locus that harbors the neurofibromin gene. Neurofibromin is a homolog of the p21 Ras-dependent GTPase-activating proteins and is thought to function in a tumor suppressor mode through regulation of Ras-dependent signaling activity.

Von Hippel-Lindau Disease

Von Hippel-Lindau disease is a heritable autosomal dominant disorder characterized by retinal and cerebellar hemangioblastomas, renal cell carcinoma, islet cell tumors, pheochromocytomas, and renal, pancreatic, and epididymal cysts. The presence of pheochromocytomas and most of the islet cell tumors is confined to the type 2 variant of the disease, which accounts for 25% to 35% of affected kindreds. The genetic lesion has been localized to 3p25. The von Hippel-Lindau protein, which is normally encoded by this locus, participates in the formation of a multiprotein complex involved in the regulation of hypoxia-induced gene transcriptional activity, fibronectin matrix assembly, and ubiquitin ligases.

REFERENCES

MEN1

Gibril F, Schumann M, Pace A, Jensen RT. Multiple endocrine neoplasia type 1 and Zollinger Ellison syndrome. *Medicine*. 2004;83:43. [PMID: 12946484]

Glascock MJ, Carty SE. Multiple endocrine neoplasia type1: fresh perspective on clinical features and penetrance. *Surg Oncol*. 2002;11:143. [PMID: 12356510]

Hughes CM, Rozenblatt-Rosen O, Milne TA, et al. Menin associates with a trithorax family histone methyltransferase complex and with the Hoxc8 locus. *Mol Cell*. 2004;13:567. [PMID: 14992727]

Lakhani VT, You YN, Wells SA. The multiple endocrine neoplasia syndromes. *Ann Rev Med*. 2007;58:253. [PMID: 17037976]

Lemos MC, Thakker RV. Multiple endocrine neoplasia type 1 (MEN1): analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mut*. 2008;29:22. [PMID: 17879353]

Matkar S, Thiel A, Hua K. Menin: a scaffold protein that controls gene expression and signaling. *Trends Endocrinol Metab*. 2013;38:394. [PMID: 23850066]

Milne TA, Hughes CM, Lloyd R, et al. Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc Natl Acad Sci (USA)*. 2005;102:749. [PMID: 15640349]

Pieterman CRC, Conemans EB, Dreijerink KMA, et al. Thoracic and duodeno-pancreatic neuroendocrine tumors in multiple endocrine neoplasia type 1: natural history and function of menin in tumorigenesis. *Endocrine Related Cancer*. 2014;21:R121. [PMID: 24389729]

Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*. 2012;97:2990. [PMID: 22723327]

Triponez F, Dosseh D, Goudet P, et al. Epidemiology data on 108 MEN1 patients from the GTE with isolated non-functioning tumors of the pancreas. *Ann Surg*. 2006;243:265. [PMID: 16432361]

Trouillas J, Labat-Moleur F, Sturm N, et al. Pituitary tumors and hyperplasia in multiple endocrine neoplasia type 1 syndrome (MEN1): a case control study in a series of 77 patients versus 2509 non-MEN1 patients. *Am J Surg Path*. 2008;32:534. [PMID: 18300794]

Yates CJ, Newey PJ, Thakker RV. Challenges and controversies in management of pancreatic neuroendocrine tumours in patients with MEN1. *Lancet Diabetes Endocrinol*. 2015;3:895. [PMID: 26165399]

MEN2

Brandi ML, Gagel RF, Angeli A, et al. Consensus guidelines for diagnosis and therapy of MEN type 1 and type 2. *J Clin Endocrinol Metab*. 2001;86:5658. [PMID: 11739416]

Elisei R, Romei C, Cosci B, et al. RET genetic screening in patients with medullary thyroid cancer and their relatives: experience with 807 individuals at one center. *J Clin Endocrinol Metab*. 2007;92:4725. [PMID: 17895320]

Machens A, Holzhausen HJ, Thanh PN, et al. Malignant progression from C-cell hyperplasia to medullary thyroid carcinoma in 167 carriers of the Ret germ-line mutations. *Surgery*. 2003;134:425. [PMID: 14555929]

Moline J, Eng C. Multiple endocrine neoplasia, type 2: an overview. *Genetics Med*. 2011;13:755. [PMID: 21552134]

Punales MKC, daRocha AP, Meotti C, et al. Clinical and oncological features of children and young adults with multiple endocrine neoplasia type 2A. *Thyroid*. 2008;18:1. [PMID: 18991485]

Runeberg-Roos P, Saarma M. Neurotrophic factor receptor RET: structure, cell biology and inherited disease. *Ann Med*. 2007;39:572. [PMID: 17934909]

Thosani S, Ayala-Ramierez M, Palmer L, et al. The characterization of pheochromocytoma and its impact on overall survival in multiple endocrine neoplasia type 2. *J Clin Endocrinol Metab*. 2013;98:E1813. [PMID:24030942]

Verga U, Fugazzola L, Cambiaghi S, et al. Frequent association between MEN 2A and cutaneous lichen amyloidosis. *Clin Endocrinol*. 2003;59:156. [PMID: 12864791]

Wells SA, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567. [PMID: 25810047]

Wells SA, Pacini F, Robinson BG, Santoro M. Multiple endocrine neoplasia type 2 and familial medullary thyroid carcinoma: an update. *J Clin Endocrinol Metab*. 2013;98:3149. [PMID: 23744408]

Other Disorders

Espiard S, Bertherat J. Carney Complex. *Front Hormone Res*. 2013;41:50. [PMID: 23652670]

Schernthaner-Reiter M, Trivellin G, Stratakis CA. MEN1, MEN4 and Carney Complex: pathology and molecular genetics. *Neuroendocrinology*. 2015;103:18. [PMID: 25592387]

Spiegel AM, Weinstein LS. Inherited diseases involving G proteins and G protein-coupled receptors. *Annu Rev Med*. 2004;55:27. [PMID: 14746508]

Transgender Endocrinology

Stephen M. Rosenthal, MD and Wylie C. Hembree, MD

5α-RD-2	5α-reductase-2		**ES**	Endocrine Society
17β-HSD-3	17β-hydroxysteroid dehydrogenase-3		**FSH**	Follicle stimulating hormone
AR	Androgen receptor		**FTM**	Female-to male
BMD	Bone mineral density		**GID**	Gender Identity Disorder
CAH	Congenital Adrenal Hyperplasia		**GnRH**	Gonadotropin releasing hormone
CAIS	Complete androgen insensitivity syndrome		**HPG**	Hypothalamic-pituitary-gonadal
DSD	Disorder of Sex Development		**LH**	Luteinizing hormone
DSM	Diagnostic and Statistical Manual of Mental Disorders		**MTF**	Male-to female
			PGR	Progesterone receptor
DXA	Dual energy x-ray absorptiometry		**SOC**	Standards of Care
ER	Estrogen receptor		**WPATH**	World Professional Association for Transgender Health

PART I: ENDOCRINE MANAGEMENT OF TRANSGENDER YOUTH

Introduction

Gender nonconforming/transgender youth, in increasing numbers, are seeking medical services to enable the development of physical characteristics consistent with their experienced gender. These services including use of agents to block endogenous puberty at Tanner stage 2 with subsequent use of cross-sex hormones are based on longitudinal studies demonstrating that those individuals who were first identified as gender dysphoric in early or middle childhood and continue to meet the mental health criteria for being transgender at early puberty are likely to be transgender as adults. The goals of **Part I** of this chapter are to review terms and definitions applicable to gender nonconforming youth (and adults), endocrine, genetic and neuroanatomical studies that shed light on the biologic underpinnings of gender identity and to review the natural history of transgenderism, current clinical practice guidelines for transgender youth, and limitations and challenges to optimal care.

Terms and Definitions

"Gender identity" refers to a person's inner sense of self as male or female (and is not always binary). In contrast, a person's "sex" refers to the physical attributes that characterize biologic maleness or femaleness (eg, the genitalia). Most children, at birth, are assigned a "sex of rearing" (some mental health providers prefer the term "gender of rearing") based on genital anatomy. With that assignment comes societal expectations of gender roles, behaviors, and expressions, which are sometimes referred to as a child's "assigned (or natal) gender." In contrast to "sex (or gender) of rearing," "gender identity" can only be assumed, and not, in fact, known until an individual achieves a particular level of psychological development and self-awareness. "Cisgender" refers to an individual for whom gender identity and physical sex characteristics are in alignment.

"Gender nonconforming" is a term sometimes used when gender identity is not consistent with the gender assumed based on birth sex assignment. As defined by the American Psychiatric Association, "transgender" refers to a person who transiently or persistently identifies with a gender different from their "natal gender." The World Professional Association for Transgender Health

(WPATH) Standards of Care (SOC) describes "transgender" (adjective) as "a diverse group of individuals who cross or transcend culturally defined categories of gender" and "transsexual" (adjective) as "individuals who seek to change or have changed their primary and/or secondary sex characteristics through feminizing or masculinizing medical interventions (hormones and/or surgery), typically accompanied by a permanent change in gender role." "Transsexual" has also been used to describe a person who identifies as a member of the gender opposite to that assigned at birth, but who has not necessarily sought medical and/or surgical interventions.

"Transgender," defined more narrowly, refers to an individual whose gender identity is the "opposite" of their "assigned" or "natal gender." A person with a female gender identity and male assigned sex would be referred to as a "transgender girl/woman," "transfemale," or MTF (male to female); a person with a male gender identity and female assigned sex would be referred to as a "transgender boy/man," "transmale," or FTM.

The terms "gender queer," "gender variant," "gender expansive," "gender creative," or "gender independent" are increasingly used to describe individuals who state that their gender identities are neither male nor female. Some of these individuals accept their assigned genders but not the cultural expectations for those genders. "Gender fluid" refers to an individual whose gender identity changes over time or is not a fixed binary. In this context, it is not clear whether core gender identity is, in fact, changing, or whether gender fluidity reflects an ongoing process of self-awareness.

"Gender behavior" is not equivalent to gender identity, and the majority of youth with gender non-conforming behavior will not turn out to have a transgender identity. In the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) 5, "Gender dysphoria" refers to clinically significant distress related to the incongruence between one's "affirmed" or "experienced gender" and one's "assigned (or natal) gender." This term was replaced "Gender Identity Disorder" (GID) found in the earlier DSM IV. The DSM 5 has distinct diagnostic criteria for "gender dysphoria in children" and "gender dysphoria in adolescents and adults" [specifying with or without a disorder of sex development (DSD)]. Replacing "disorder" with "dysphoria" depathologizes the transgender identity, and instead, focuses on dysphoria as the clinical problem. It should be noted that the majority of patients with "gender dysphoria" do not have a DSD. Whether in children or adolescents, a core feature of "gender dysphoria" is "a marked incongruence between one's experienced/expressed gender and assigned gender" of at least 6 months' duration.

"Sexual identity" (or "sexual orientation") is often confused with "gender identity." While the former refers to whom one is sexually attracted, the latter refers to whom one "is" as male, female, or somewhere on the gender continuum. As with cisgender individuals, gender nonconforming/transgender individuals may have a heterosexual, homosexual, or bisexual orientation.

Prevalence of Transgenderism in Youth

While prevalence data for younger transgender adolescents are lacking, multidisciplinary clinics for transgender youth and adolescents in Europe and North America have seen a steadily increasing demand for services in recent years, with a ratio of phenotypic males to females close to 1:1.

Mental Health Concerns and Impact of Family Support

Life-threatening behaviors represent a significant risk for transgender youth and adolescents. A retrospective cohort study of 180 transgender individuals aged 12 to 29 years seen in a Boston clinic between 2002 and 2011 were compared to an equal number of cis-gender, age-matched controls. In comparison to controls, transgender youth/young adults had a two- to threefold increased risk of anxiety, depression, suicidal ideation, suicide attempts, and self-harm without lethal intent. The impact of the degree of parental support on mental health outcomes was reported in a cohort of transgender youth and young adults age 16 to 24 years (n = 84) from Ontario, Canada: Satisfaction with life and self-esteem were significantly greater in transgender youth whose parents were "very supportive" versus those whose parents were "somewhat to not at all supportive." In addition, depression and suicide attempts were significantly decreased in transgender youth whose parents were supportive in comparison to those whose parents were not supportive. Even with supportive parents, transgender youth still have a significant risk for depression, perhaps in part from their experience of transphobia from members of their communities and feelings of rejection and social isolation.

Biologic Underpinnings of Gender Identity

Numerous studies from a variety of biomedical disciplines—endocrine, genetic, and neuroanatomical—have begun to shed light on the biologic underpinnings of gender identity. Results of these studies support the concept that gender identity is not simply a psychosocial construct, but likely reflects a complex interplay of biologic, environmental, and cultural factors.

While the majority of transgender patients do not have a DSD, studies of gender identity outcome within the endocrine discipline have been principally carried out in patients with a variety of DSDs, primarily exploring the role of prenatal and postnatal androgens in gender identity development. For example, studies in 46, XX patients with "classical" congenital adrenal hyperplasia (CAH) caused by mutations in the *CYP21A2* gene resulting in 21-hydroxylase deficiency and varying degrees of genital masculinization, demonstrate a greater than expected number of patients with gender dysphoria, "atypical gender identity," or who were transgender. In a meta-analysis of 250 patients with 46, XX 21-hydroxylase deficiency leading to virilizing CAH and raised female, nearly 95% of patients reported a female gender identity without gender dysphoria, while 5.2% reported either a male gender identity or gender dysphoria. Of note, there was no apparent correlation with the degree of genital masculinization and gender identity outcome. A study of adult women with classical 21-hydroxylase deficiency demonstrated a relationship between severity of disease and gender identity outcome: Of 42 patients

with the salt-wasting form, 3 (7.1%) had either gender dysphoria or had changed gender to male; no gender dysphoria was seen in less severely affected individuals. These and other studies demonstrate that the majority of 46, XX patients with virilizing CAH from 21-hydroylase deficiency appear to have a female gender identity. However, the finding that 5.2% to 11.6% of such patients have gender dysphoria, an "atypical gender identity" or are transgender would appear to be much more common than expected based on the reported prevalence of FTM transsexualism, implying that there is some role for prenatal/postnatal androgens in gender identity outcome. It is noteworthy that in 46, XX individuals with virilizing CAH from 21-hydroylase deficiency, prenatal androgens are more likely to affect gender behavior and sexual orientation than gender identity.

The potential effects of prenatal and postnatal androgen exposure on gender identity outcome and "gender role change" have also been explored in other hormonal and non-hormonal DSDs. For example, in 5α-reductase-2 deficiency (5α-RD-2), among 46 XY individuals raised female, a "gender role change from female to male" (typically after puberty) was reported in 56% to 63% of the patients. Gender role changes from female to male were also reported in 39% to 64% of 46 XY individuals raised female with 17β-hydroxysteroid dehydrogenase-3 deficiency (17β-HSD-3). In the largest series of patients with 5α-RD-2 and 17β-HSD-3 reviewed by Cohen-Kettenis and colleagues, individuals undergoing gender role change from female to male had intact testes, implying a potential role of prenatal as well as postnatal androgen exposure in gender identity outcome.

Gender identity outcome has also been studied in patients with a variety of "non-hormonal" DSDs, including cloacal exstrophy, penile ablation, and penile agenesis. A study of patients with 46 XY cloacal exstrophy reported that the majority (8 of 14) who had undergone neonatal sex reassignment to female (castration) subsequently declared a male gender identity, and that 2 patients raised as males remained male. A literature review of individuals with 46 XY cloacal exstrophy found that of 51 patients assigned female, the majority (65%) were living as female while 14% were living as female but with possible gender dysphoria, and approximately 22% were living as male. In addition, of 16 males with penile agenesis assigned female at birth and of 7 males with penile ablation reassigned to female in infancy or early childhood, the majority were living as female.

These studies of gender identity outcome in hormonal and non-hormonal DSDs, taken together, indicate that gender identity is not solely dependent on prenatal and postnatal androgen exposure; however, the occurrence of gender identity change (in comparison to the "natal gender") at a rate significantly higher than would be expected in the general population supports some role of prenatal and possibly postnatal androgens in gender identity development. Potential limitations of all these survey/questionnaire-based studies to assess gender identity include a person's degree of self-awareness and one's willingness to disclose this information in the study context.

While the above-noted studies support some role of androgens in male gender identity, a case report in a 46, XY individual with complete androgen insensitivity syndrome (CAIS) challenges the concept that androgen receptor (AR) signaling is required for male gender identity development. Individuals with CAIS typically have unambiguous female genitalia and a female gender identity; however, an individual with CAIS, associated with an unambiguous female phenotype and an *AR* gene mutation resulting in a premature stop codon, was found to have a male gender identity (26). As acknowledged by the authors of that study, this patient may have had a post-zygotic *de novo AR* gene mutation (the mother's DNA was not available) and could theoretically have brain mosaicism (with a normal AR in the brain).

With respect to genetics and gender identity, heritability of transsexualism has been suggested primarily from studies describing concordance of transsexualism in monozygotic twin pairs. A study of survey responses from parents of twins (96 monozygotic pairs and 61 dizygotic pairs, ages 7-14 years) demonstrated a 2.3% prevalence of clinically significant GID (based on DSM-IV criteria). These authors reported that heritability accounted for 62% of the variance for GID in their twin sample. A more recent study supporting a role for genetic factors in gender identity outcome demonstrated a 39.1% concordance for GID (based on DSM-IV criteria) in 23 monozygotic female and male twin pairs, with no concordance in 21 same-sex dizygotic female and male twin pairs or in 7 opposite-sex twin pairs. However, with respect to specific candidate genes [estrogen receptor α (*ER*α) and β (*ER*β), androgen receptor (*AR*), aromatase, (*CYP19*), progesterone receptor (*PGR*), and the *CYP17* gene], association studies of polymorphisms with transsexualism have been inconsistent and lacking strong statistical significance.

With respect to neuroanatomical considerations and gender identity, it has been known for some time that some brain structures are sexually dimorphic, while one component of the hypothalamus has been reported to be sexual orientation-dimorphic. Numerous studies in transgender adults, carried out prior to any cross-sex hormone treatment, have provided evidence indicating that some sexually dimorphic structures appear to align more closely with a person's gender identity than with their physical sex characteristics. A trend towards similar findings has also been reported in MRI studies of gray matter of gender dysphoric youth.

Transgender Youth: Natural History

Once puberty has begun, the majority of gender dysphoric prepubertal youth will no longer meet the mental health criteria for gender dysphoria. While gender fluidity may be a contributing factor, the lack of persistence of gender dysphoria in the majority of gender dysphoric prepubertal youth likely reflects the heterogeneous nature of this group. Many such individuals will turn out to be homosexual (natal males, in particular) rather than transgender. Investigators have attempted to identify factors that predict gender dysphoria "persisters" versus "desisters." Persisters reported relatively greater degrees of gender dysphoria and were more likely to have experienced social transition (to their affirmed gender) during childhood. In addition, persisters believed they "were" the other sex, while desisters "wished they were" the other sex. The limitations in prediction of persistence, coupled with the

observation that most gender dysphoric children will not become transgender adolescents or adults, have led some investigators and/or practitioners to promote efforts to encourage gender dysphoric children to accept their natal gender. In contrast, a model of care that affirms a child's gender expression is thought to have a more optimal mental health outcome. Long-term outcomes studies are needed to resolve these differences in approach to the care of gender dysphoric prepubertal youth.

Clinical Practice Guidelines for Transgender Youth

Persistence of gender dysphoria following initiation of puberty implies a very high likelihood that such individuals will be transgender as adults; in fact, the emergence or worsening of gender dysphoria with onset of puberty is thought to have significant diagnostic value in the determination of being transgender. With the availability of gender-affirming medical interventions (as detailed later), however, awareness of such treatment options may positively impact and, thus, limit the degree of dysphoria that might otherwise accompany the onset of puberty in a transgender individual. It is therefore essential that gender dysphoric youth undergo a thorough psychological diagnostic evaluation by a qualified mental health provider. An important role of the mental health/gender specialist is not only to determine the presence or absence of gender dysphoria, but also to evaluate for the presence of other mental health concerns. For example, while there is an increased association of gender dysphoria and autism spectrum disorder, the majority of gender dysphoric children and adolescents do not have an underlying severe psychiatric illness.

Management of Early Pubertal Transgender Youth

Pubertal suppression with a gonadotropin releasing hormone (GnRH) agonist (classically used for the treatment of precocious puberty) in early/mid-pubertal gender dysphoric adolescents was first described in studies from the Netherlands. This treatment is fully reversible and allows additional time for gender exploration without the pressure of ongoing pubertal development. The physical changes of puberty, once completed, cannot be reversed (by means other than surgical or, for voice, other than by voice training)—for example, low voice, Adam's apple, and facial features in phenotypic males and breast development in phenotypic females. Preventing pubertal development that does not match a person's gender identity can theoretically lead to decreased distress and can ultimately enable the individual to more easily "blend" in society as an adult. Subsequently, if the individual continues to identify as transgender, cross-sex hormones can be added while continuing GnRH agonist suppression of endogenous puberty, enabling the individual to experience only the physical changes of puberty that match the person's experienced/affirmed gender identity. The Endocrine Society (ES) guidelines and WPATH SOC endorse the use of pubertal blockers (using GnRH agonists) at Tanner 2/3 in individuals experiencing a significant increase in gender dysphoria with onset of puberty. While age-specific

guidelines for subsequent interventions are not delineated in the WPATH SOC, the ES guidelines suggest that cross-sex hormones can be initiated at about the age of 16 years (the legal age for medical decision-making in some countries), while surgical procedures (with the exception of mastectomy) should be deferred until the individual is at least 18 years of age. Delaying cross-sex hormone treatment until age 16 could not only be detrimental to bone health, but, by keeping someone in a prepubertal state until this age, would isolate the individual further from age-matched peers, with potentially negative consequences for emotional well-being. Gender centers at our institution and elsewhere are therefore studying the impact of cross-sex hormone treatment initiation at 14 years of age (which approximates the upper end of the age range for normal pubertal onset in natal males and 1 year beyond the upper end of the age range in natal females). In this group, sex steroids are increased gradually over the course of 2 to 3 years.

The above-noted ES and WPATH recommendations are supported by limited available outcomes data. A prospective follow-up study from the Netherlands assessed 70 gender dysphoric adolescents (33 MTF, 37 FTM), average age 13.65 years (range 11.1-17 years) at initial assessment, who were Tanner Stage 2 to 3, had "life-long gender dysphoria" that increased with puberty, had stable psychological function, and were supported by their families. These adolescents were studied at the initiation of GnRH agonist treatment and approximately 2 years later, just before starting cross-sex hormones. While gender dysphoria persisted, depressive symptoms decreased, general mental health functioning improved, no subjects withdrew from pubertal suppression, and all went on to cross-sex hormone treatment. A subsequent report from this initial cohort in the Netherlands assessed mental health outcomes in 55 transgender adolescents/young adults (22 MTF, 33 FTM) at 3 time points: before the start of GnRH agonist treatment, at initiation of cross-sex hormones (average 16.7 years at start of treatment), and at least 1 year after "gender reassignment surgery" (average age 20.7 years). Despite a decrease in depression and an improvement in general mental health functioning, gender dysphoria persisted during the period of pubertal suppression (as noted in earlier reports); however, following cross-sex hormone treatment and "gender reassignment surgery," gender dysphoria was resolved, and psychological functioning steadily improved. Well-being was found to be similar to or better than that in age-matched young adults from the general population, and none of the study participants regretted treatment. This study represents the first long-term follow-up of patients managed according to currently existing clinical practice guidelines for transgender youth and underscores the benefit of the multidisciplinary approach to care.

The primary risks of pubertal suppression in gender dysphoric youth treated with GnRH agonists include adverse effects on bone mineralization, compromised fertility, and unknown effects on brain development.

Bone mineral density (BMD) studies of the lumbar spine, femoral neck, and total body were initially carried out in a small number of gender dysphoric adolescents during 2 years of GnRH agonist alone and for an additional 2 years of combination treatment with GnRH agonist and cross-sex hormones. BMD (g/cm^2) did not change significantly during

treatment with GnRH agonist alone, though z-scores decreased; during combination therapy with GnRH agonist and cross-sex hormones, absolute BMD increased, as did z-scores.

A 6-year longitudinal, observational study assessed BMD in 34 transgender adolescents (15 MTF, 19 FTM) who had received GnRH agonist beginning at an average age of 14.9 to 15 years (individuals were mid-late pubertal at study onset by testicular volume or breast stage), had initiation of cross-sex hormones at 16.4 to 16.6 years, followed by gonadectomy with discontinuation of GnRH agonist at a minimum age of 18 years. Over the 6-year observation period, areal BMD z-scores decreased significantly in MTF individuals with a trend for a decrease in FTM individuals, suggesting either a delay in attainment of peak bone mass, or an attenuation of peak bone mass, itself. As acknowledged by the authors, potential study limitations included a relatively small number of subjects, the fact that individuals were already late pubertal at the time of GnRH agonist initiation, the possibility that relatively low doses of cross-sex hormones were used during the initial period of that phase of treatment, and that information was not available regarding dietary calcium intake, vitamin D levels, and weight bearing exercise, all of which can influence BMD. It would seem important, particularly during GnRH agonist treatment, to ensure adequate intake of calcium and vitamin D, to encourage weight-bearing exercise, and to monitor levels of 25-OH vitamin D routinely.

Any use of pubertal blockers and cross-sex hormones in transgender youth should include an informed consent process and a discussion about implications for fertility. Transgender adolescents may wish to preserve fertility, which may be otherwise compromised if puberty is suppressed at an early stage and the patient completes phenotypic transition with the use of cross-sex hormones. While full *in vitro* maturation of germ cells has not yet been achieved in humans, promising studies have been carried out in mice.

Regarding potential adverse effects of GnRH agonists (and cross-sex hormones) on the brain, a recent study investigated whether GnRH agonist treatment of gender dysphoric adolescents would impair executive functioning (reasoning, problem solving, etc.), thought to reflect prefrontal brain activation, and to have significant development during puberty. No significant detrimental effects of GnRH agonist treatment on executive functioning were observed.

A report from Vancouver, Canada, assessed potential adverse effects of GnRH agonist treatment in transgender youth. Relatively few adverse effects were observed: Of 27 patients receiving GnRH agonist, one developed a sterile abscess at the injection site, one developed transient leg pains and headaches, and one developed significant weight gain.

Management of Late Pubertal Transgender Youth

It is not unusual for some transgender adolescents to first come to medical attention when they are late pubertal (Tanner 4/5). FTM individuals can be treated with testosterone alone while MTF individuals are optimally treated with estrogen with concurrent use of an agent that blocks testosterone secretion and/or action.

GnRH agonists can be used to suppress the hypothalamic-pituitary-gonadal axis, potentially enabling use of lower doses of cross-sex hormones to induce phenotypic transition to the affirmed gender, thereby decreasing potential toxicities associated with cross-sex hormone treatment. In this older cohort, cross-sex hormone regimens may be increased to full replacement doses over a shorter interval than used for the younger cohort that had been initially treated with pubertal blockers at Tanner 2/3. While GnRH agonists are the preferred option for pubertal suppression in both the early and late pubertal individuals, this treatment is costly and often inaccessible. Options for pubertal suppression and cross-sex hormone treatment are listed in Table 23–1.

TABLE 23–1 Hormonal interventions for transgender adolescents (all currently off-label for gender non-conforming/transgender youth).

A. Inhibitors of gonadal sex steroid secretion or action

1. GnRH analogues: Inhibition of the hypothalamic-pituitary-gonadal (HPG) axis (FTM and MTF)

   a. Leuprolide acetate IM (1- or 3-month preparations) or SC (1-, 3-, 4-, or 6-month preparations) at dose sufficient to suppress pituitary gonadotropins and gonadal sex steroids

   b. Histrelin acetate 50 mg SC implant (once-yearly dosing, though may have longer effectiveness)

   c. Other options: goserelin acetate SC implant (4- or 12-week preparations); nafarelin acetate intranasal (multiple daily doses) also available, but no reported use in this population

2. Alternative approaches

   a. Medroxyprogesterone acetate orally (up to 40 mg/d) or IM (150 mg q 3 months): Inhibition of HPG axis and direct inhibition of gonadal steroidogenesis (FTM and MTF)

   b. Spironolactone (25-50 mg/d with gradual increase to 100-300 mg/d orally, divided into BID dosing): Inhibition of testosterone synthesis and action (MTF)

   c. Cyproterone acetate (gradual increase up to 100 mg/d orally; not available in United States): Inhibition of testosterone synthesis and action (MTF)

   d. Finasteride (2.5-5 mg/d orally): Inhibition of Type II 5α-reductase, blocking conversion of testosterone to 5α-dihydrotestosterone (MTF)

B. Cross-sex hormones

1. MTF: Estrogen: 17 β-estradiol

   a. Transdermal: twice weekly patches [6.25 μg (achieved by cutting a 25 mcg patch) with gradual increase to full adult dose (typically, 100-200 μg patch)] Oral/sublingual: daily (0.25 mg with gradual increase to full adult dose of 2-6 mg/d)

   b. Parenteral IM (synthetic esters of 17 β-estradiol): estradiol valerate (5-20 mg increasing up to 30-40 mg q 2 weeks) or estradiol cypionate (2-10 mg q 1 week)

2. FTM: Testosterone

   a. Parenteral IM or SC (synthetic esters of testosterone): testosterone cypionate or enanthate (12.5 mg q week or 25 mg q 2 weeks, with gradual increase to 50-100 mg q week or 100-200 mg q 2 weeks)

   b. Transdermal (consider once full adult testosterone dose has been achieved parenterally): patch (2-6 mg/d) or 1.62% gel (20.25-81 mg/d)

Abbreviations: MTF, Male-to-Female; FTM, Female-to-Male.

Reproduced with permission from Rosenthal SM. Approach to the patient: transgender youth: endocrine considerations. *J Clin Endocrinol Metab.* 2014 Dec;99(12): 4379-4389.

With respect to estrogen treatment, 17 β-estradiol (transdermal, oral, or parenteral) is preferred to conjugated (eg, premarin) or synthetic estrogens (eg, ethinyl estradiol) given that conjugated and synthetic estrogen levels cannot be monitored in the serum and that ethinyl estradiol (in comparison to 17 β-estradiol) is associated with an increased risk for venous thromboembolic disease and death from cardiovascular causes. MTF individuals treated with estrogen may have impaired insulin sensitivity and hyperprolactinemia. Principal risks associated with testosterone treatment in FTM individuals include cystic acne, polycythemia, hypertension, an atherogenic lipid profile, and possible decreased insulin sensitivity. Table 23–2 summarizes surveillance recommendations for desired as well as adverse effects during treatment with pubertal blockers alone and in combination with cross-sex hormones, adapted from the current ES guidelines.

Some Tanner 4/5-transgender adolescents present for medical services before 16 years of age. As with the group treated with GnRH agonists at early puberty, we and others are studying the consequences of cross-sex hormone treatment at 14 years of age. Some gender dysphoric youth first come to medical attention when they are Tanner 4/5, but less than 14 years of age. These

TABLE 23–2   Monitoring during pubertal suppression and during cross-sex hormone treatment.*

A. Pubertal suppression	
Measure	Frequency
1. Physical examination: height, weight, Tanner Staging	T 0 & q 3 mo
2. Hormonal studies: ultrasensitive LH, FSH, estradiol/testosterone	T 0 & q 3 mo
3. Metabolic: Ca, phos, alk phos, 25-OH Vitamin D (see also ref. 3)	T 0 & q 1 yr
4. Bone mineral density: DXA	T 0 & q 1 yr
5. Bone age	T 0 & q 1 yr
B. Cross-sex hormone treatment in previously suppressed patients or in late pubertal patients not previously suppressed	
Measure	Frequency
1. Physical examination: height, weight, Tanner Staging, BP (for FTM, in particular); monitor for adverse reactions	T 0 & q 3 mo**
2. Hormonal studies: ultrasensitive LH, FSH, estradiol/testosterone	T 0 & q 3 mo**
If MTF: also monitor prolactin	T 0 & q 1 yr
3. Metabolic: Ca, phos, alk phos, 25-OH vitamin D, complete blood count, renal & liver function, fasting lipids, glucose, insulin, glycosylated hemoglobin	T 0 & q 3 mo**
If MTF on spironolactone: serum electrolytes (potassium)	T 0 & q 3 mo**
4. Bone mineral density: DXA (if puberty previously suppressed)	T 0 & q 1 yr***
5. Bone age (if puberty previously suppressed)	T 0 & q 1 yr***

*Modified from Hembree WC et al., 2009; **q 3-12 mo after 1st yr; ***Until puberty is completed.

Abbreviations: MTF, Male-to-Female; FTM, Female-to-Male.

Data from Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, et al. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2009 Sep;94(9):3132-3154.

individuals would be candidates for pubertal blockers (eg, to stop menses in an FTM adolescent), but without supportive outcomes data, they would not currently be candidates for cross-sex hormone use under most circumstances.

Areas of Uncertainty/Barriers to Care/and Priorities for Research

Areas of uncertainty and barriers to state-of-the art practice limit the ability to provide optimal health care to gender dysphoric/transgender youth. Currently, there are only limited safety and efficacy studies, with virtually no published data on the use of pubertal blockers in gender dysphoric individuals less than 12 years of age or cross-sex hormones in transgender youth less than 16 years of age. In addition, randomized controlled trials for hormonal interventions in gender dysphoric youth have not been considered feasible or ethical. Current clinical practice guidelines are based on best available evidence, with significant reliance on expert opinion. The Institute of Medicine of the National Academies in the United States has endorsed the need for prospective, longitudinal safety, and efficacy studies of medical interventions in gender non-conforming/transgender youth. Challenges to implementation of current clinical practice guidelines in the United States include the fact that pubertal blockers and cross-sex hormone treatments are off-label for gender dysphoric youth, are expensive, and coverage is often denied by insurance companies. In addition, while an increasing number of clinical programs have emerged in recent years, there are many geographic regions in which such services do not exist, limiting access to care and often requiring patients and families to travel long distances. Furthermore, lack of training of providers and prejudice and misunderstanding on the part of family, community, and medical and mental health professionals may limit access to optimal care. A report surveying the memberships of the Pediatric Endocrine Society and the Society for Adolescent Health and Medicine explored providers' clinical experiences, comfort, and confidence with and barriers to providing care to transgender youth. Principal barriers to provision of transgender-related care included lack of the following: training, exposure to transgender patients, available qualified mental health providers, and insurance reimbursement.

Endocrine Management of Transgender Youth: Conclusions

Significant advances in our understanding of gender nonconforming/transgender youth have been achieved, though many gaps in knowledge remain. Compelling studies have demonstrated that gender identity is not simply a psychosocial construct, but likely reflects a complex interplay of biologic, environmental, and cultural factors. The replacement of "disorder" with "dysphoria" in DSM-5 removes the connotation that a transgender identity, itself, is pathologic, and instead, focuses on dysphoria as the clinical concern. A landmark study from the Netherlands indicates that mental health comorbidities in gender-dysphoric youth significantly diminish or resolve when such

individuals are subject to a gender-affirming model of care, optimally delivered in a multidisciplinary clinical setting. Further prospective studies focused on long-term safety and efficacy are necessary to optimize medical and mental health care for transgender youth.

PART II: ENDOCRINE MANAGEMENT OF TRANSGENDER ADULTS

Introduction

The age adulthood begins is variably defined at age 18 to 21 years, at which time persons can independently give informed consent for invasive procedures. Most persons this age who have persistent gender dysphoria are treated with adult protocols. The natural history of gender dysphoria in adults has not been well documented although when most adults present for hormone treatment they describe the occurrence of symptoms beginning during childhood. The goals of **Part II** of this chapter are to review adult presentation of gender dysphoria, endocrine status in adults with gender dysphoria, strategies for sex steroid treatment, potential adverse effects of estradiol and testosterone treatment, genital and breast surgery options, reproductive decisions, the role of voice therapy, and considerations for care of elder transgender individuals.

Adult Presentation of Gender Dysphoria

Adults with gender dysphoria variably present to mental health professionals, to physicians experienced in endocrine treatment, to surgeons, or to clinics with all of these types of providers. Appropriate clinical management may require all of the above. The first step is confirmation of the diagnosis, gender dysphoria, assessment of readiness of the person to begin treatment and evaluation of risk factors for medical treatment (Table 23–3). Rarely, adults with psychological and/or developmental problems may present with gender problems but are not confirmed to exhibit gender dysphoria. The second step is exploration of appropriate hormonal protocols, including the associated risks, benefits, and contraindications. Third, hormone treatment is initiated using varying protocols suitable for transgender women (MTF) or transgender men (FTM) (Table 23–4). Consultation with surgeons who specialize in genital surgery occurs after at least 1 year of cross-sex hormone treatment. Breast and facial surgery are commonly explored.

In clinical practice, persons with gender dysphoria who seek treatment with sex steroids, genital surgery and social transition to a new gender may present at any time during adulthood. Clinical histories usually describe gender-related conflicts with their natal sex, especially during puberty. Many have married, become parents, and lived as a person of their natal sex, a man or a woman. Conflicts between natal sex and gender often occasion presentation to a mental health professional, although the diagnosis of gender dysphoria may not be apparent. Most give a history of incomplete resolution of childhood symptoms despite living in their natal sex as adults. Exploration of the components

TABLE 23–3 Hormone protocols for transgender adults.

A. Male-to-female (MTF) protocols

a. Phase 1: Blocking testosterone production
(May be given initially to reduce testosterone levels or simultaneously with estrogen)

i. Spironolactone	100-200 mg/d
ii. Cyproterone acetate	50-100 mg/d
iii. GnRH analog	3.76 mg/mo
(Gonadotropin Releasing Hormone Agonist)	

b. Phase 2: Estradiol administration

i. Oral estradiol	2.0-5.0 mg daily
ii. Transdermal E2 patch	0.1-0.4 2-3× weekly*
iii. Parenteral E2 valerate or cypionate	2-10 mg/wk

B. Female-to-male (FTM) protocols

a. Phase 1: Blocking ovulation/estradiol production
(Note: Testosterone injections alone at early follicular phase may block ovulation)

i. Medroxyprogesterone acetate	30-40 mg/d
ii. Cyproterone acetate	50-100 mg/d
iii. GnRH analog	3.75-15 mg q 1-3 mo

b. Phase 2: Testosterone administration
(Note: Phase 1 medications may be given before, simultaneously or not at all with testosterone)

i. Transdermal	
1. Testosterone Gel 1%	2.5-10 mg/d
2. Testosterone transdermal patch	2.5-10 mg/q 2-3 d
ii. Parenteral	
1. Testosterone enanthate/cypionate	50-100 mg q 2 wks
	30-60 mg q wk
2. Testosterone undecanoate	1000 mg q 8-12 wks

*Recommended for MTF over 40 years old

**Synthetic estradiol esters not recommended

TABLE 23–4 Transgender medical evaluation.

1. Review gender dysphoria history
   a. If present, review mental health professional recommendation

2. Medical and mental health history
   a. Assess risk factors to sex steroids
      i. Transgender women (MTF)
         1. Thromboembolic disease/history
         2. Hyperprolactinemia
         3. Cerebrovascular disease
         4. Severe migraine headaches
         5. Breast cancer
      ii. Transgender men (FTM)
         1. Erythrocytosis
         2. Severe liver dysfunction
         3. Hypertension
         4. Breast or uterine cancer
   b. Family history
      i. Reproductive tract cancer
      ii. Diagnosis of mental health illness
      iii. Osteoporosis/history of bone fracture

3. Medications and allergies

4. Alcohol use, smoking, addiction history, hospitalizations

5. Physical examination
   a. Breast examination
   b. Reproductive tract examination

of gender change includes discussions of hormone treatment as well as surgical options. New onset of gender dysphoria as an adult is unusual.

Endocrine Considerations and Management

Endocrine evaluation of adults with gender dysphoria usually discovers normal development of the body that reflects natal sex changes as well as, after puberty, body changes produced by appearance of their natal sex hormones. However, amongst some persons with a DSD, CAH, AIS, and 5α-RD, there is more frequent occurrence of gender dysphoria and gender change than in the general population. Studies of the relationship of early exposure to androgens may explain sexual and social behavior, but they do not clarify patient-initiated gender choice. However, circulating levels of testosterone, metabolism to estrogens, and impact upon bone development in natal males are determined, in large part, genetically. This may account for the variability observed in bone responses during sex steroid treatment.

The design of the sex steroid protocols used is based upon, first, the treatment that will most appropriately and completely suppress natal sex steroids, either before or at the same time exogenous sex steroids are given (see Table 23–4). Second, the protocol should utilize the most effective available method by which sex steroids of the person's new gender can be administered. For example, some FTM protocols may use testosterone alone, thereby suppressing ovulation, preventing uterine bleeding, achieving effective virilization, while avoiding adverse effects. On the other hand, MTF protocols using physiologic replacement doses of estradiol alone may not sufficiently suppress endogenous testosterone levels and, thereby, result in insufficient feminization. In such cases, it may be necessary to use additional agents, initially alone or as a combination with estrogens, that will effectively suppress endogenous testosterone levels to those of natal women. Progestins have been used frequently to suppress endogenous sex steroids, as have GnRH analogs. The doses of exogenous sex steroids required to suppress endogenous sex steroids may vary with the individual, the dose used, the method of administration and their natal sex.

The goal of testosterone treatment in transgender men (FTM) with regular menses is to suppress ovulation and, thereby, maintain low estradiol levels. Thus, the time to initiate testosterone treatment in protocols for transgender men (FTM) is important. Treatment should begin during or immediately following menses when estradiol and progesterone levels are lowest, when there is little stimulation of the endometrium and there is no follicular development (estrogen levels are in the male normal range). Levels of free testosterone should be maintained above the mid-range of the adult natal male in order to suppress follicular development and uterine bleeding. Testosterone may be given by intramuscular injection or by transdermal application of testosterone gel. Transdermal patches do not permit easy dose adjustment. Total and free testosterone, total estradiol, and LH/FSH levels should be measured at 3-month intervals initially to ensure ovarian suppression and to avoid adverse effects (Table 23–5).

TABLE 23–5   Monitoring sex steroid treatment in adults male-to-female (MTF).

1. Signs of feminization
   a. Evaluate 3-4 times year 1 and 1-2 times years 2-4
   b. Breast growth
   c. Loss of terminal hair growth
   d. Body fat redistribution
   e. Decrease testicular size
2. Assess hormone levels
   a. Evaluate 3-4 times year 1 and 1-2 times year 2-4
   b. Estradiol: 100-200 pg/dL
   c. Testosterone: <50 ng/dL
   d. Prolactin: Pretreatment, annual
   e. LH/FSH: if testosterone levels >100 ng/dL
3. Spironolactone use
   a. Electrolytes (potassium): Baseline, every 2-3 months × 2 and after increased dose × 2
4. Bone mineral density assessment
   a. Baseline if risk factors present
   b. Age 60 years old
   c. Non-compliant sex steroid use
   d. Cessation of sex steroids
5. Routine breast, colon, and prostate cancer screening

Female-to-male (FTM)
1. Signs of virilization:
   a. Evaluate 3-4 times year 1 and 1-2 times years 2-4
   b. Cessation of menses
   c. Facial/body hair growth
   d. Increase muscle mass
   e. Deepening of voice
   f. Clitoral enlargement
2. Assess hormone levels
   a. Measure testosterone every 3 months until levels remain in physiologic range of natal adult male;
   b. Testosterone enanthate/cypionate: demonstrate peak and trough levels within physiologic range of natal adult male × 2;
   c. Transdermal testosterone: assess after 2 weeks of use, measure more than 2 hours after application;
   d. Testosterone undecanoate: measure testosterone prior to next injection; if <400 ng/dL, increase dose/interval;
3. Obtain weight, CBC, blood pressure, lipids every 3-4 months year 1 and 1-2 times yearly thereafter;
4. Annual PAP if cervix present;
5. If mastectomy, regular examination of periareolar tissue; if no mastectomy, mammograms as recommended by the American Cancer Society.

When transgender men are given high doses of testosterone, the biologically active free testosterone is determined by the clearance rate of testosterone, the binding to SHBG and the aromatization to estradiol. Variation in each of the above will determine the appropriate testosterone dose needed for each person. At the onset of puberty in natal males and during testosterone replacement in hypogonadal men, high aromatase levels may result in high estradiol levels and transient gynecomastia. In adult natal women who are transgender men (FTM), aromatase activity is higher than adult men. Thus, when a full adult male dose of testosterone is given to an FTM, high estradiol levels may result. Little is known about aromatization of testosterone in the endometrium and its potential for causing uterine bleeding, although aromatase inhibitors have been used successfully to treat

endometrial spotting. Published protocols using testosterone alone for transgender men (FTM) report cessation of ovulation and little uterine bleeding. However, initial suppression of ovulation with progestins or GnRH agonist may be necessary. In addition to the effect of changes in aromatase levels, alterations in testosterone clearance and binding by sex hormone binding globulin (SHBG) may alter the testosterone dose required. Monitoring of blood pressure and hemoglobin levels will prevent adverse effects; while estradiol and gonadotropin levels will monitor ovarian suppression (see Table 23–5).

Estradiol treatment of transgender women (MTF) requires suppression of endogenous testosterone levels from 300-1000 ng/dL to less than 50 ng/dL. Oral, transdermal and intramuscular preparations are available. Estradiol levels in natal, ovulatory women average 100-200 pg/mL and should be the goal for transgender women. However, estrogen doses that achieve this range of estradiol in MTF may fail to suppress endogenous testosterone levels to less than 50 ng/dL. Natal women whose testosterone levels exceed 50 ng/dL may exhibit hirsutism or signs of mild virilization, despite good estradiol levels. Thus, optimal suppression of endogenous testosterone by exogenous estradiol usually requires use of a second medication. Compounds with progestin activity (spironolactone, medroxyprogesterone, and cyproterone acetate) are most commonly used. GnRH analogs also effectively reduce testosterone levels to less than 50 ng/dL. Medications with 5α-reductase inhibitory activity may reduce balding in transgender men and women. However, their use in MTF protocols is not advised due to their failure to suppress endogenous testosterone levels in transgender women and their adverse effects.

Protocols that effectively suppress endogenous sex steroids and maintain adult levels of the new exogenous sex steroid should have periodic assessment during the first 1 to 2 years of treatment (see Table 23–5). Variations in transdermal absorption occur over time, as does metabolic clearance, and SHBG levels. Most of these changes occur within the first 6 to 12 months of treatment. Stable sex steroid levels achieve gender-specific body changes that begin within 6 months and stabilize within 2 to 5 years. Effects of estrogen include changes in body fat, muscles and skin, decreases in sexual function, libido, breast growth, minor decreases in testicular size, decreased body hair growth, and slowing of male baldness. Effects of testosterone include acne, body hair growth, scalp hair loss, changes in muscle mass and body fat, voice deepening, genital changes, and clitoral enlargement. Increasing sex steroid doses does not change the time course of these changes in adults.

Surveillance for Potential Adverse Effects of Hormonal Treatment

Transgender men and women should be screened for heart disease, hypertension, metabolic syndrome, lipid disorders, diabetes, and breast and prostate cancer. The frequency of this monitoring varies from 1 to 4 times during the first year to annually after 3 years. Elevated prolactin levels and prolactinomas, as well as excess cases of thrombophlebitis, have been reported in transgender women. Data suggest that ethinyl estradiol increases the risk of vascular disease in transgender women. Breast development is variable in MTF persons, is considered inadequate by most transgender women and is not improved by use of progestins. There is a high request rate for mammoplasty. Low bone mass has been observed in transgender women despite physiologic levels of estradiol. Cardiovascular disease is higher in MTF persons treated with estrogen than in testosterone treated FTM persons, reversing the trend observed in natal men and women. Surveillance of laboratory tests that monitor general health should occur at baseline and be monitored periodically. Normal ranges should be interpreted with caution since studies have shown that the "normal range" for transgender men and women may not be coincident with those of natal men and women.

Surgical Considerations

Transgender men and women may elect to have both major and minor surgical procedures during gender transition. Transgender women select facial treatment over chest and genital procedures whilst transgender men opt for chest/breast procedures. MTF persons, especially those 20 to 30 years old, may plan genital surgery that creates a neovagina. A variety of techniques have been developed, each with advantages and disadvantages. Clitoroplasty is widely used and results in good sexual sensitivity; creation of labia minora and labia majora utilizes a variety of techniques with variable success. Breast development during estradiol treatment occurs during a 1- to 2-year period, and outcomes are variable. Breast augmentation is requested in approximately 50% of transgender women. Standard breast cancer screening is required after 2 years of estrogen treatment. For FTM persons, phalloplasty is available in a greater number of centers, although metoidioplasty is selected more frequently, preserving sensitivity and facilitating urination while standing. Mastectomy in FTM persons may leave residual subareolar breast tissue for cosmetic reasons. Close surveillance for cancer screening is required. It is unlikely that screening for endometrial and ovarian cancer occurs regularly in FTM persons and, thus, hysterectomy/ovariectomy should be considered.

Reproductive Options

Both transgender women and transgender men may elect to have children. A variety of options are available, although the number of transgender persons electing pregnancy is low. The impact of gender transition upon future reproductive options should be discussed with all transgender persons who are seeking hormonal or surgical transition. In transgender women, sperm cryopreservation prior to estrogen treatment is available at modest cost. Although 51% of MTF persons indicated they would elect sperm cryopreservation, published reports of pregnancy using cryopreserved sperm from transgender women are few. Testis morphology in adult men treated with estradiol suggests that residual Leydig cells, Sertoli cells and spermatogonia remain, although no studies have demonstrated recovery of spermatogenesis following cessation of estradiol. Transgender men who have transitioned with testosterone and maintained reproductive organs have successfully elected pregnancy. Oocyte, ovarian, and embryo cryopreservation are available prior to initiation of testosterone treatment in

transgender men, although the chances of achieving live births from these tissues are small. In most states and countries, transgender families may also elect adoption.

Voice Therapy

Increasingly, transgender persons seek advice and treatment from speech language pathologists, although one survey indicates that many felt improperly prepared and that specific training courses are limited. Half of transgender persons surveyed were unaware of available services. Correlation of speech patterns with body movement also aids identity as female. One chart review of 25 persons undergoing feminizing hormone treatment found that 80% successfully present as female 100% of the time. Happiness of MTF with their voice after voice training correlates more so with perceived femininity and quality of life than with achieved pitch. MTF adolescents may also benefit from voice and communication therapy. Surgical techniques are available for MTF transgender persons. However, comparison of the outcomes of speech therapy alone, surgical techniques alone, and surgery followed by speech therapy has not been studied. Of the three approaches for FTM, there are no outcome comparison studies. In contrast to MTF persons, intonation shows little correlation with male identity. Long-term follow-up of androgen effects upon voice in transgender men do not differ from those observed in biological males. Hematocrit and CAG repeat lengths in the androgen receptor (an indication of androgen sensitivity) correlate directly with vocal androgen sensitivity and reflect the extent of voice deepening when treated with testosterone. Difficulties in pitch-lowering occur only in 10% of FTM persons treated long term with testosterone and may be related to decreased androgen sensitivity. Medical providers of transgender persons, especially MTF, should be aware that speech language pathologists indicate that working with transgender persons is within their scope of practice.

Aging and Transgender Care

Increasingly, transgender persons have reached the age of 50 years. Many were referred at a younger age to mental health professionals, providers skilled in sex steroid treatment, and specialized surgeons who provided guidance throughout their gender transition. However, specialized needs for long-term follow-up medical care needed during aging are less well known amongst generalists who serve as primary care physicians for transgender persons. Maintenance of secondary sex characteristics in both MTF and FTM persons is a central concern of most transgender persons. In addition, concerns about cancer risk are prominent. Regular breast and genital examinations, as well as sexually transmitted disease screening, occur infrequently in transgender men whereas these examinations are frequent in transgender women. No data are available for prostate and PSA surveillance in MTF persons.

Considerations for continuing sex steroid treatment (estrogen and testosterone) are similar for transgender men and transgender women to those for natal men and natal women, respectively. In MTF and FTM who had gonadectomy, LH and FSH levels can be used to monitor maintenance of prescribed doses of sex steroids. There are no data that estradiol treatment causes prolactin-secreting pituitary adenomas, although estrogen treatment may increase prolactin levels when adenomas are present. Small increases in prolactin may occur during estradiol treatment in the absence of adenomas, although these increases usually remain within the normal range and are stable during estrogen treatment. Thus, screening of prolactin levels is recommended before starting estrogen treatment and should continue periodically in MTF persons who continue estradiol treatment. Use of transdermal estradiol is preferable in older age, especially for those with heart disease, hypertension, bone disease, and risk of thrombophlebitis. As in older natal women, dose reduction may be considered for risk reduction. Regular prostate examinations are recommended although cases of prostate cancer or hypertrophy are rare. The reverse trend in cardiovascular risk factor surveillance should continue in both MTF and FTM persons.

FTM persons who elect hysterectomy may continue testosterone treatment without concerns about aromatization to estradiol. As suggested above, ovariectomy should be performed when hysterectomy is elected. As in natal men, normal testosterone levels in transgender older men exert no adverse effects upon cardiovascular disease, diabetes, hypertension or metabolic syndrome. Subareolar breast tissue in FTM should be monitored for rare cases of cancer in residual tissue following mastectomy. Bone density in older men is usually maintained by testosterone and the estradiol produced by aromatization of testosterone. Periodic BMD assessments should be done in both transgender men and transgender women, especially when sex steroid treatment is stopped.

Endocrine Management of Transgender Adults: Conclusions

Treatment of adult transgender men and transgender women with physiologic replacement doses of testosterone and estradiol, respectively, facilitates the psychological, metabolic and physical changes that resolve gender dysphoria and enables a successful lifestyle transition to that of their appropriate gender. The principles of effective treatment are to suppress the natal sex steroids and to replace them with sex steroid levels appropriate for the transgender person. Guidelines for safe and effective treatment, noting both risks and benefits, are available to physicians providing care during the transition and during long-term care. Health-care coverage and protection against discrimination increasingly enable life fulfillment for transgender persons.

REFERENCES

Transgender Youth

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA; American Psychiatric Association Publishing; 2013.

Berenbaum SA, Bailey JM. Effects on gender identity of prenatal androgens and genital appearance: evidence from girls with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2003;88:1102. [PMID: 12629091]

Cohen-Kettenis PT. Gender change in 46, XY persons with 5-alpha-reductase-2 deficiency and 17-beta-hydroxysteroid dehydrogenase-3 deficiency. *Arch Sex Behav*. 2005;34:399. [PMID: 16010463]

Coleman E, Bockting W, Botzer M, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgenderism*. 2011;13:165.

de Vries AL, McGuire JK, Steensma TD, Wagenaar ECF, Doreleijers TAH, Cohen-Kettenis PT. Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*. 2014;134:696. [PMID: 25201798]

de Vries AL, Noens IL, Cohen-Kettenis PT, van Berckelaer-Onnes IA, Doreleijers TA. Autism spectrum disorders in gender dysphoric children and adolescents. *J Autism Dev Disord*. 2010;40:930. [PMID: 20094764]

de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *J Sex Med*. 2011;8:2276. [PMID: 20646177]

Delemarre-van de Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspects. *Eur J Endocrinol*. 2006;155:S131.

Dessens AB, Slijper FM, Drop SL. Gender dysphoria and gender change in chromosomal females with congenital adrenal hyperplasia. *Arch Sex Behav*. 2005;34:389. [PMID: 16010462]

Gooren LJ. Clinical practice. Care of transsexual persons. *N Engl J Med*. 2011;364:1251. [PMID: 21449788]

Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, et al. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2009;94:3132. [PMID: 19509099]

Heylens G, De Cuypere G, Zucker KJ, et al. Gender identity disorder in twins: a review of the case report literature. *J Sex Med*. 2012;9:751. [PMID: 22146048]

Hidalgo MA, Ehrensaft D, Tishelman AC, et al. The gender affirmative model: what we know and what we aim to learn. *Hum Dev*. 2013;56:285.

Hoekzema E, Schagen SEE, Kreukels BPC, et al. Regional volumes and spatial volumetric distribution of gray matter in the gender dysphoric brain. *Psychoneuroendocrinology*. 2015;55:59. [PMID: 25720349]

Imperato-McGinley J, Peterson RE, Gautier T, Sturla E. Androgens and the evolution of male-gender identity among male pseudohermaphrodites with 5-alpha-reductase deficiency. *N Engl J Med*. 1979;300:1233. [PMID: 431680]

Khatchadourian K, Amed S, Metzger DL. Clinical management of youth with gender dysphoria in Vancouver. *J Pediatr*. 2014;164:906. [PMID: 24315505]

Klink D, Caris M, Heijboer A, van Trotsenburg M, Rotteveel J. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *J Clin Endocrinol Metab*. 2015;100:E270. [PMID: 25427144]

Meyer-Bahlburg HF. Gender identity outcome in female-raised 46, XY persons with penile agenesis, cloacal exstrophy of the bladder, or penile ablation. *Arch Sex Behav*. 2005;34:423. [PMID: 16010465]

Meyer-Bahlburg HF, Dolezal C, Baker SW, Ehrhardt AA, New MI. Gender development in women with congenital adrenal hyperplasia as a function of disorder severity. *Arch Sex Behav*. 2006;35:667. [PMID: 16902816]

Reiner WG, Gearhart JP. Discordant sexual identity in some genetic males with cloacal exstrophy assigned to female sex at birth. *N Engl J Med*. 2004;350:333. [PMID: 14736925]

Reisner SL, Vetters R, Leclerc M, et al. Mental health of transgender youth in care at an adolescent urban community health center: a matched retrospective cohort study. *J Adolesc Health*. 2015;56:274. [PMID: 25577670]

Rosenthal SM. Approach to the patient: transgender youth: endocrine considerations. *J Clin Endocrinol Metab*. 2014;99:4379. [PMID: 25140398]

Rosler A, Silverstein S, Abeliovich D. A (R80Q) mutation in 17 beta-hydroxysteroid dehydrogenase type 3 gene among Arabs of Israel is associated with pseudohermaphroditism in males and normal asymptomatic females. *J Clin Endocrinol Metab*. 1996;81:1827. [PMID: 8626842]

Sherer I, Baum J, Ehrensaft D, Rosenthal SM. Affirming gender: caring for gender atypical children and adolescents. *Contemporary Pediatrics*. 2015;32:16.

Sherer I, Rosenthal SM, Ehrensaft D, Baum J. Child and adolescent gender center: a multidisciplinary collaboration to improve the lives of gender nonconforming children and teens. *Pediatr Rev*. 2012;33:273. [PMID: 22659258]

Spack NP, Edwards-Leeper L, Feldman HA, et al. Children and adolescents with gender identity disorder referred to a pediatric medical center. *Pediatrics*. 2012;129:418. [PMID: 22351896]

Staphorsius AS, Kreukels BPC, Cohen-Kettenis PT, et al. Puberty suppression and executive functioning: an fMRI-study in adolescents with gender dysphoria. *Psychoneuroendocrinology*. 2015;56:190. [PMID: 25837854]

Steensma TD, McGuire JK, Kreukels BP, Beekman AJ, Cohen-Kettenis PT. Factors associated with desistence and persistence of childhood gender dysphoria: a quantitative follow-up study. *J Am Acad Child Adolesc Psychiatry*. 2013;52:582. [PMID: 23702447]

Travers R, Bauer G, Pyne J, et al. Impacts of strong parental support for trans youth: a report prepared for Children's Aid Society of Toronto and Delisle Youth Services. *Trans Pulse*. 2012;1.

T'Sjoen G, De Cuypere G, Monstrey S, et al. Male gender identity in complete androgen insensitivity syndrome. *Arch Sex Behav*. 2011;40:635. [PMID: 20358272]

Vance SR, Ehrensaft D, Rosenthal SM. Psychological and medical care of gender nonconforming youth. *Pediatrics*. 2014;134:1184. [PMID: 25404716]

Vance SR, Halpern-Felsher BL, Rosenthal SM. Health care providers' comfort with and barriers to care of transgender youth. *J Adolesc Health*. 2015;56:251. [PMID: 25620310]

Zucker KJ, Wood H, Singh D, Bradley SJ. A developmental, biopsychosocial model for the treatment of children with gender identity disorder. *J Homosex*. 2012;59:369. [PMID: 22455326]

Transgender Adults

Asscherman H, T'Sjoen G, Lemaire A, et al. Venous thromboembolism as a complication of cross-sex hormone treatment of male-to-female subjects: A review. *Andrologia*. 2014;46:791. [PMID: 23944849]

Berenbaum SA, Meyer-Bahlburg HFL. Gender development and sexuality in disorders of sex development. *Horm Metab Res*. 2015;47:361. [PMID: 25853895]

Cosyns M, Van Borse J, Wierckx K, et al. Voice in female-to-male transsexual persons after long term androgen therapy. *Laryngoscope*. 2014;124:1409. [PMID: 24155064]

Ettner R. Care of the elderly transgender patient. *Curr Opin Endocrinol Diabetes Obes*. 2013;30:580. [PMID: 24468762]

Gooren L, Lips P. Conjectures concerning cross-sex hormone treatment of aging transsexual persons. *J Sex Med*. 2014;11:2012. [PMID: 24775178]

Gooren LJ, Wierckx K, Giltay EJ. Cardiovascular disease in transsexual persons treated with cross-sex hormones: reversal of the traditional sex difference in cardiovascular disease pattern. *Eur J Endocrinol*. 2014;170:809. [PMID: 24626414]

Hancock A, Colton L, Douglas F. Intonation and gender perception: applications for transgender speakers. *J Voice*. 2014;28:203. [PMID: 24094799]

Hancock AB, Garabedian LM. Transgender voice and communication treatment: a retrospective chart review of 25 cases. *Int J Lang Commun Disord*. 2013;48:54. [PMID 23317384]

Hancock AB, Krissinger J, Owen K. Voice perceptions and quality of life of transgender people. *J Voice*. 2011;25:553. [PMID: 21051199]

Lapauw B, Taes Y, Simoens S, et al. Body composition and areal bone parameters in male-to-female transsexual persons. *Bone*. 2008;43:1016. [PMID: 18835591]

Mueller A, Binder H, Cupisti S, et al. Effects on the male endocrine system of long-term treatment with gonadotropin-releasing hormone agonists and estrogens in male-to-female transsexuals. *Horm Metab Res*. 2006;38:183. [PMID: 16673210]

Meuller A, Haeberie L, Zoliver H, et al. Effects of intramuscular testosterone undecanoate on body composition and bone mineral density in female-to-male transsexuals. *J Sex Med*. 2010;7:3190. [PMID: 20584125]

Meuller A, Zollver H, Kronawiller D, et al. Body composition and bone mineral density in male-to-female transsexuals during cross-sex hormone therapy using gonadotropin-releasing hormone agonist. *Exp Clin Endocrinol Diabetes*. 2011;119:95. [PMID: 20625973]

Meyer-Bahlburg HFL. Sex steroids and variants of gender identity. *Endocrinol Metab Clin N Am*. 2013;42:435. [PMID: 24011879]

Perrone AM, Cerpolini S, Maria Safli NC, et al. Effect of long-term testosterone administration on the endometrium of female-to-male (FtM) transsexuals. *J Sex Med*. 2008;6:3193. [PMID: 19570144]

Simpson E, Santen RJ. Celebrating 75 years of oestradiol. *J Mol Endocrinol*. 2015;55:T1. [PMID: 25438567]

Stricker R, Eberhart R, Chavailler MC, et al. Establishment of detailed reference values for luteinizing hormone, follicle stimulating hormone, estradiol, and progesterone during different phases of the menstrual cycle on the Abbott ARCHITECT analyzer. *Clin Chem Lab Med*. 2006;44:883. [PMID: 16776638]

Travison TG, Zhuang WV, Lunetta KL, et al. The heritability of circulating testosterone, estradiol, estrone, and SHBG concentrations in men: the Framingham Study. *Clin Endocrinol (Oxf)*. 2014;80:1. [PMID: 23746309]

Wierckx K, Gooren L, T'Sjoen G. Clinical review: breast development in trans women receiving cross-sex hormones. *J Sex Med*. 2014;11:1240. [PMID: 24618412]

Wierckx K, Mueller S, Weyers S, et al. Long-term evaluation of cross-sex hormone treatment in transsexual persons. *J Sex Med*. 2012;10:2641. [PMID: 22906135]

Wierckx K, Stuyver I, Weyers S, et al. Sperm freezing in transsexual women. *Arch Sex Behav*. 2012;42:1069. [PMID: 22968492]

Wierckx K, Van Caenegem E, Pennings G, et al. Reproductive wish in transsexual men. *Hum Reprod*. 2012;27:483. [PMID: 22128292]

Wierckx K, Van Caenegem E, Schreiner T, et al. Cross-sex hormone therapy in trans persons is safe and effective at short-term follow-up: results from the European Network for the investigation of gender incongruence. *J Sex Med*. 2014;11:1999. [PMID: 24828032]

Wierckx K, Van de Peer F, Verhaeghe E, et al. Short- and long-term clinical effects of testosterone treatment in trans men. *J Sex Med*. 2014;11:222. [PMID: 24828032]

Wroblewski P, Gustafsson J, Selvaggi G. Sex reassignment surgery for transsexuals. *Curr Opin Endocrinol Diabetes Obes*. 2013;20:570. [PMID: 24468760]

Endocrine Emergencies

David G. Gardner, MD, MS

ACTH	Adrenocorticotropic hormone	**L-T$_4$**	Levothyroxine
ADH	Antidiuretic hormone (vasopressin)	**MRI**	Magnetic resonance imaging
APACHE	Acute physiology and chronic health evaluation	**PTH**	Parathyroid hormone
		PTHrP	Parathyroid hormone–related protein
CT	Computed tomography	**SIADH**	Syndrome of inappropriate antidiuretic hormone secretion
DDAVP	1-deamino-8-D arginine vasopressin		
DKA	Diabetic ketoacidosis	**T$_3$**	Triiodothyronine
FT$_4$	Free thyroxine	**TPP**	Thyrotoxic periodic paralysis
GFR	Glomerular filtration rate	**TSH**	Thyroid-stimulating hormone
IL	Interleukin		

Acute or chronic failure of an endocrine gland can occasionally result in catastrophic illness and even death. Thus, it is important to recognize and appropriately manage these endocrine emergencies. This chapter will discuss crises involving the thyroid, anterior pituitary, or adrenal glands; diabetes mellitus; and abnormalities in calcium, sodium, and water balance. Except where indicated, management recommendations are provided for adult patients. Studies in the general area of endocrine emergencies have been limited in size and number. In many instances, recommendations offered in this chapter are based on published expert opinion rather than scientific evidence.

MYXEDEMA COMA

Clinical Setting

Myxedema coma is the end stage of untreated or inadequately treated hypothyroidism. It has an estimated incidence of 0.22 per million per year. The clinical picture is often that of an elderly obese female, presenting in midwinter with increased lethargy, somnolence, and confusion. The presentation is one of severe hypothyroidism, with or without coma (the term myxedema coma may, therefore, be a misnomer). The history from the patient may be inadequate, but the family may report that the patient has had

thyroid surgery or radioiodine treatment in the past or that the patient has previously been receiving thyroid hormone therapy. Myxedema coma is most frequently associated with discontinuation of thyroid hormone therapy. It presents less frequently as the first manifestation of hypothyroidism. It may be precipitated by an illness such as a cerebrovascular accident, myocardial infarction, or an infection such as a urinary tract infection or pneumonia. Other precipitating factors include gastrointestinal hemorrhage; acute trauma; excessive hydration; or administration of a sedative, narcotic, or potent diuretic drug.

Diagnosis

The physical findings are not specific. The patient may be semicomatose or comatose with dry, coarse skin, hoarse voice, thin scalp and eyebrow hair, possibly a scar on the neck, and slow reflex relaxation time. There is marked hypothermia, with body temperature sometimes falling to as low as 24°C (75°F), particularly in the winter months. It is important to be alert to the presence of complicating factors such as pneumonia, urinary tract infection, ileus, anemia, hypoglycemia, or seizures. Fever may be masked by coexistent hypothermia. Often there are pericardial, pleural, or peritoneal effusions. The key laboratory tests are a low free thyroxine (FT$_4$) and elevated thyroid-stimulating hormone (TSH). The TSH elevation may be less than predicted due to the presence of

euthyroid sick syndrome or glucocorticoid or dopamine therapy. TSH may also be low when central hypothyroidism is present—myxedema coma has a pituitary or hypothalamic etiology approximately 5% of the time. Central hypothyroidism is characterized by low FT_4 and low or normal TSH. Pituitary insufficiency can be confirmed with a low serum cortisol, impaired response to the cosyntropin stimulation test, and/or low follicle-stimulating hormone and luteinizing hormone. It is essential to check blood gases, electrolytes, creatinine, and an electrocardiogram in evaluating pulmonary, renal, cardiac, and central nervous system status. Myxedema may cause prolongation of the Q-T interval leading to torsade de pointes. The combination of hypothermia, hyponatremia, and hypercapnia should increase suspicion of myxedema coma. It may be necessary to differentiate myxedema coma from the euthyroid sick syndrome associated with coma due to other causes. These patients may present with a low triiodothyronine (T_3), normal or low TSH, but the FT_4 (by dialysis) is normal.

Myxedema coma is a complex problem involving a number of organ systems. The pathogenesis is presented in Figure 24–1. The decrease in serum T_4 results in a lowering of intracellular T_3. This can directly affect central nervous system function with altered mental status. The decrease in intracellular T_3 causes decreased thermogenesis, resulting in hypothermia, which in turn causes decreased central nervous system sensitivity to hypercapnia and hypoxia. The resulting respiratory insufficiency induces cerebral anoxia and coma. While coma, per se, is the dominant clinical manifestation, disorientation, depression, and psychosis with hallucinations (ie, myxedema madness) are often seen. At the same time, the decreased intracellular T_3 results in decreased cardiac inotropism and chronotropism, decreased sensitivity to adrenergic stimuli, decreased cardiac output, and generalized vasoconstriction. This leads to a low-output state which, if untreated, culminates in decreased blood pressure and eventually shock and death. Finally, there is a change in fluid balance with increased water retention due to impaired renal perfusion and syndrome of inappropriate antidiuretic hormone secretion (SIADH), as well as increased vascular permeability. These changes result in effusions and hyponatremia, which in turn contributes to the coma. There is also an increased risk of bleeding (eg, gastrointestinal bleeding) due to an acquired coagulopathy reflecting deficiency of a number of clotting factors (von Willebrand factor as well as factors V, VII, VIII, IX, and X).

Management

Management of myxedema coma involves much more than simply replacing T_4 (Table 24–1). The patient is severely ill and usually is admitted to the intensive care unit (ICU) for intubation and ventilatory support. Oral medications may be poorly absorbed (due to gastric atony or ileus), and medications should be given intravenously if possible. A loading dose of 300 to 400 μg of L-T_4 intravenously is given initially to saturate T_4 binding sites in plasma-binding proteins. The patient is then maintained on 50 to 100 μg of L-T_4 intravenously daily. In addition, small doses of T_3 (eg, 10 μg intravenously every 6-8 hours) may be given over the first 48 hours, but this is usually not necessary, and it may contribute to untoward cardiovascular events. Its use should be restricted to younger patients without history of ischemic heart disease or cardiac arrhythmias. Water restriction is necessary to correct the hyponatremia, and intravenous glucose counteracts the tendency
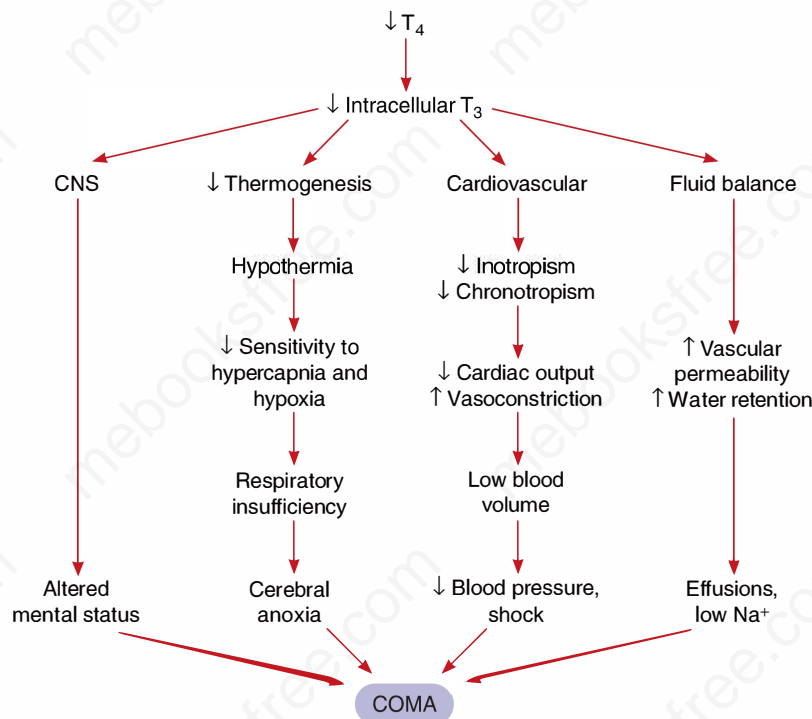
FIGURE 24–1 Pathogenesis of myxedema coma (CNS, central nervous system). (See text for details.)

TABLE 24–1 Management of myxedema coma.

(1) Admit to ICU for ventilatory support and for intravenous medications.
(2) Parenteral thyroxine: Give a loading dose of 300-400 μg IV, then 50-100 μg IV daily. (May also give liothyronine sodium, 10 μg IV every 8 h for the first 48 h if necessary.)
(3) Electrolytes: Water restriction for hyponatremia. Avoid fluid overload.
(4) Limit sedation. Appropriate reduction in drug dosage.
(5) Glucocorticoids: Controversial but necessary in hypopituitarism or polyglandular failure. Dosage: Hydrocortisone sodium phosphate or sodium succinate, 50-100 mg every 6 h initially and tapered downward over 1 wk. (If initial serum cortisol was >30 μg/dL, corticosteroids are unnecessary.)
(6) Hypothermia: Do not externally rewarm.

TABLE 24–2 Thyroid storm—precipitating factors.

Withdrawal of antithyroid drugs
Severe infection
Diabetic ketoacidosis
Myocardial infarction
Cerebrovascular accident
Cardiac failure
Surgery
Parturition
Trauma (eg, hip fracture)
Radioiodine (rare)
Drug reaction
Iodinated contrast medium

to hypoglycemia. It is essential to treat the precipitating illness (eg, pneumonia or urinary tract infection). The use of hydrocortisone is prudent because patients may have hypopituitarism or autoimmune polyglandular failure. Glucocorticoids are generally given in high doses until baseline cortisol or the results of rapid adrenocorticotropic hormone (ACTH) testing are available. If the initial serum cortisol is more than 30 μg/dL, steroid support is probably unnecessary. However, if serum cortisol is less than 30 μg/dL, hydrocortisone should be given intravenously in a dosage of 50 to 100 mg every 6 hours for the first 48 hours and the dose then tapered over the next 5 to 7 days while the pituitary-adrenal axis undergoes formal testing. Once myxedema coma is suspected in a hypothermic patient, external rewarming should be avoided, because this may cause redistribution of blood flow to subcutaneous tissues and cardiovascular collapse.

Prior to the recognition of the need for intravenous L-T_4 and for respiratory support, the mortality from myxedema coma was about 80%. Currently, the mortality is about 20% and is mostly due to the underlying or precipitating illness. Predictors of acute mortality include level of consciousness, lower Glasgow scores, and higher (acute physiology and chronic health evaluation) APACHE II scores, all indicative of the severity of the illness at presentation. Higher mortality is also associated with increased age, cardiac complications, and high-dose thyroid hormone replacement when this has been used (≥500 μg/d of L-T_4 or ≥75 μg/d of L-T_3). The latter presumably reflects the increased metabolic demand attendant to high dose replacement in the setting of limited physiologic reserve. Persistent hypothermia and bradycardia, despite therapy, are associated with a poor prognosis.

THYROID STORM

Clinical Setting

Thyroid storm, or thyrotoxic crisis, is an acute, life-threatening exacerbation of thyrotoxicosis. It accounts for 1% to 2% of hospital admissions for thyrotoxicosis. It may occur in a patient with a history of Graves disease who has discontinued antithyroid medication or in a patient with previously undiagnosed hyperthyroidism. There are rare reports of "storm" with subacute thyroiditis or

overdose with exogenous L-T_4. The clinical picture is that of an acute onset of hyperpyrexia (with temperature >40°C [104°F]), sweating, marked tachycardia often with atrial fibrillation, nausea, vomiting, diarrhea, agitation, tremulousness, and altered mentation. The presence of jaundice is considered a poor prognostic sign. Occasionally, the presentation is apathetic without the restlessness and agitation, but with symptoms of weakness, confusion, cardiovascular and gastrointestinal dysfunction, and hyperpyrexia. Some of the factors that may precipitate thyroid storm are listed in Table 24–2.

Diagnosis

The diagnosis is largely based in the clinical findings. Serum T_4, free T_4, T_3, and free T_3 are all elevated, and TSH is suppressed. Patients present with hyperthermia, often out of proportion to the magnitude of coincident infection, diaphoresis, tachycardia (often with supraventricular tachyarrhythmias and high output heart failure), and central nervous system dysfunction. The latter most clearly distinguishes thyroid storm from severe, but compensated, hyperthyroidism. It is thought that thyroid storm represents an exacerbation of thyrotoxicosis associated with a shift of T_4 from the bound to the free compartment with an increase in free T_3 and T_4, as well as an exaggerated response to a surge of catecholamines that results from the stress of the precipitating event. The cause of death is usually cardiac arrhythmia and failure. Liver function abnormalities are often seen, as is leukocytosis, even in the absence of infection. Hypercoagulability is related to increased levels of a number of clotting factors (eg, factors II, VIII, and IX).

Management

The management of thyroid storm is summarized in Table 24–3. Patients are typically admitted to an ICU for cardiac monitoring. Initially, it is important to block further synthesis and secretion of thyroid hormone, first with antithyroid drugs and then with iodide. One may give propylthiouracil, 150 mg every 6 hours orally or rectally, or methimazole, 20 mg every 8 hours orally or rectally. A few hours after initiation of antithyroid drug therapy, iodides may be started. Traditionally, saturated solution of potassium iodide, 5 drops twice daily, has been used, but, if available, iopanoic acid in a dose of 0.5 g twice daily orally or intravenously, or iohexol 0.6 g (2 mL of Omnipaque 300) intravenously twice daily is the

TABLE 24–3   Management of thyroid storm.

Supportive care
Fluids
Oxygen
Cooling blanket
Acetaminophen
Multivitamins
If indicated, antibiotics, digoxin
Specific measures
Propranolol, 40-80 mg orally every 6 h
Propylthiouracil, 150 mg every 6 h, or methimazole, 20 mg every 8 h; may be administered per rectum if oral route is unavailable
Saturated solution of potassium iodide, 5 drops (250 mg) orally twice daily; or iopanoic acid, 0.5 g IV or orally twice daily; or iohexol, 0.6 g (2 mL of Omnipaque 300) IV twice daily
Dexamethasone, 2 mg every 6 h
Cholestyramine or colestipol, 20-30 g/d

treatment of choice. These drugs not only inhibit thyroid hormone synthesis but also block the conversion of T_4 to T_3, lowering the thyroid hormone level in the blood. Unfortunately, these agents are not readily available in the United States. In patients who are allergic to iodine, lithium carbonate may be used to block thyroid hormone release. It is administered initially at 300 mg every 6 hours with subsequent monitoring to maintain lithium levels in the therapeutic range (0.8-1.2 mEq/L). Additional specific therapy includes β-adrenergic blockade with propranolol, 40 to 80 mg orally every 6 hours or 0.5 to 1 mg intravenously over 10 minutes every 3 hours, with continuous cardiac monitoring. In patients with asthma, the $β_1$-selective antagonist esmolol can be given as a bolus dose (0.25-0.5 mg/kg over 10 minutes) followed by continuous intravenous infusion (0.05-0.1 mg/kg/min). The half-life of glucocorticoids is markedly reduced in severe thyrotoxicosis, so that adrenal support may be very useful. Dexamethasone also reduces conversion of T_4 to T_3 (as do the beta-blockers, especially propranolol). For this purpose, dexamethasone is given in a dosage of 2 mg every 6 hours for 48 hours, followed by tapering of the dose. Cholestyramine or colestipol binds T_4 in the gut interfering with its enterohepatic circulation and may help bring the circulating level of T_4 down more quickly. Plasmapheresis and plasma exchange have been used to reduce serum FT_4 levels emergently in nonresponders, but the durability of response is limited to 24 to 48 hours, requiring a plan for definitive therapy in the short term. Emergent total thyroidectomy may be considered for those patients who fail to respond to medical therapy in a timely fashion, particularly older patients with cardiorespiratory failure. Supportive therapy includes adequate fluids, oxygen, treatment of the precipitating illness (eg, infection), management of atrial fibrillation or heart failure, if present, parenteral water-soluble vitamins, and a cooling blanket and acetaminophen for hyperpyrexia. Aspirin should be avoided, because it will displace T_4 from thyroid hormone–binding globulin, resulting in an increase in FT_4.

Therapy for thyroid storm has improved markedly, so the mortality has dropped from 100% in the 1920s to about 20% to 30% in recent series.

THYROTOXIC PERIODIC PARALYSIS

Clinical Setting

Thyrotoxic periodic paralysis (TPP) is a rare but frightening thyroid emergency. Incidence in the United States has been estimated at 0.1%-0.2% of thyrotoxic patients. The usual clinical presentation is of an Asian male (male:female ratio approximately 17:1) with symptoms of untreated hyperthyroidism who awakens at night or in the early morning hours with flaccid ascending paralysis. Typically, there is a history of vigorous exercise, alcohol abuse, and/or a large high-carbohydrate meal before retiring. There is usually no family history of periodic paralysis, but there may be a family history of autoimmune thyroid disease. Initially the patient presents with muscle cramps and stiffness. Over time this progresses to mild weakness and, at the extreme, total paralysis. The paralysis initially involves the lower extremities but progresses to the girdle muscles, followed by the upper extremities. Proximal muscle groups, typically beginning in the lower extremities, are affected to a greater extent than distal. Sensory function, facial and respiratory muscles are usually spared. Bowel and bladder function may be compromised in attacks affecting all four extremities. Deep tendon reflexes are depressed or absent. The acute episode may be complicated by cardiac arrhythmias due to the concomitant presence of hypokalemia. The illness has also been reported to occur in Native Americans, African Americans, Polynesians, Nepalese, Lebanese, and Turks and in individuals of Mexican or South American descent, but it is unusual in these latter ethnic groups. Paralysis of specific muscle groups recovers in the reverse order of appearance—usually within 36 and virtually always within 72 hours.

Diagnosis

The differential diagnosis of TPP includes familial periodic paralysis, Guillain-Barré syndrome, and acute intermittent porphyria. The diagnosis is based on the absence of a family history, the characteristic presentation, the presence of hyperthyroidism due either to Graves disease or toxic nodular goiter (other types of hyperthyroidism have been implicated as well), and usually a low serum potassium level.

The pathogenesis is summarized in Figure 24–2. Thyrotoxicosis, increased β-adrenergic activity, and an assumed genetic predisposition, perhaps involving a defective inward-rectifying potassium channel Kir2.6, which together with increased Na^+-K^+ ATPase activity leads to increased intracellular potassium concentrations. A high-carbohydrate meal with increased insulin secretion and glycogen deposition, vigorous exercise, high salt intake, and the normal nocturnal potassium flux serve to drive serum potassium levels even lower, resulting in flaccid neuromuscular paralysis. Note that there is no loss of total body potassium, merely a shift from the extracellular to the intracellular space. Consequently, aggressive potassium repletion, particularly through the parenteral route, is discouraged because this frequently results in significant *hyperkalemia* as potassium redistributes across cellular compartments following resolution of the attack. Hypophosphatemia and hypomagnesemia may be present. This is also assumed to reflect
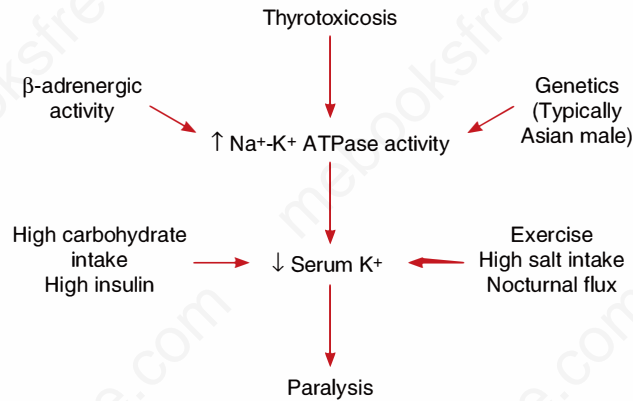
FIGURE 24–2 Pathogenesis of thyrotoxic periodic paralysis. (See text for details.)

intracellular sequestration. The electromyogram, performed while the patient is experiencing weakness, shows myopathic changes with reduced amplitude of compound muscle action potentials. These do not change in amplitude after administration of intra-arterial low-dose epinephrine (distinguishes from familial periodic paralysis). Electrocardiograms show changes associated with hypo-kalemia, tachycardia, increased QRS voltage, first-degree heart block, and, on occasion, serious ventricular arrhythmias.

Management

The management of this problem is presented in Table 24–4. Propranolol in doses of 60 mg every 6 hours blocks the β-adrenergic stimulation of Na$^+$-K$^+$ ATPase. Antithyroid drug therapy should be started immediately, even though it takes time to bring the patient into a euthyroid state. Oral potassium, if indicated, should be administered cautiously. It is particularly important to be cautious if administering intravenous potassium, which may raise total body potassium to toxic levels as the episode resolves. One should also avoid intravenous glucose, which stimulates insulin secretion and worsens hypokalemia, and β-adrenergic agonists such as isoproterenol, which promote movement of potassium into the intracellular compartment and exacerbate the problem. There is no role for potassium supplementation in pre-venting attacks. Acetazolamide, which has been shown to reduce frequency of attacks in familial periodic paralysis, may worsen attacks of TPP and should be avoided. With appropriate treat-ment, recovery is rapid, and once the thyrotoxicosis is controlled, the paralysis will not recur.

TABLE 24–4 Management of thyrotoxic periodic paralysis.

(1) Oral potassium supplement (if needed); monitor serum K$^+$
(2) Oral propranolol (60 mg every 6 h)
(3) Antithyroid drug therapy

Avoid:
IV potassium
IV glucose (ie, crystalloid fluid replacement only)
β-Adrenergic agonists (eg, isoproterenol)

AMIODARONE-INDUCED THYROTOXICOSIS

Clinical Setting

Amiodarone is a benzofuran derivative that is widely used in the treatment of cardiac arrhythmias. It contains two atoms of iodine per molecule, which represents 37.5% iodine by weight. The compound is stored in adipose tissue and has a half-life in the body of 2 to 3 months, with gradual and continuous release of iodide. The usual daily maintenance dose of amiodarone of 200 to 400 mg/d releases 6000 to 12,000 μg of iodine daily, which when compared to the normal daily requirement of about 150 μg of iodine represents an enormous iodine load.

The structure of amiodarone resembles that of T$_3$, and it is thought that part of the cardiac depressant effect of amiodarone may be due to binding to and blocking the thyroid hormone receptor in cardiac muscle. However, the effect of amiodarone on the thyroid gland is different and is due in part to a direct effect of iodine on the thyroid follicular cell—to inhibit or stimulate hor-mone synthesis—and a cytotoxic effect of amiodarone that leads to destruction of the follicular cell and release of stored hormone. Thus, the drug may induce hypothyroidism, which is easily man-aged by thyroxine replacement, or hyperthyroidism, which, because of the underlying heart disease, is much more difficult to manage and may represent a true thyroid emergency. Two mecha-nisms have been suggested to explain the development of hyper-thyroidism: (1) the high iodine level in a multinodular gland, or in the gland of a patient with latent Graves disease, or even in a previ-ously normal gland, can induce hyperthyroidism (see Chapter 7), and (2) the toxic effect of amiodarone itself may cause acute and chronic thyroiditis with release of T$_4$ and T$_3$ into the circulation and severe thyrotoxicosis. Risk factors for amiodarone-induced thyrotoxicosis include the presence of autoimmune thyroiditis and baseline elevation of TSH levels.

The patient with amiodarone-induced thyrotoxicosis may have been on the drug for months. Thyrotoxicosis typically appears 6 to 12 months after the onset of treatment. The underlying heart disease gradually worsens with increasingly frequent episodes of arrhythmia and heart failure. At the same time there may be weight loss, heat intolerance, increased nervousness, and marked muscle weakness. On physical examination, one may find non-tender nodular or diffuse thyroid enlargement, tachycardia with or without atrial fibrillation, tremor, hyperreflexia, and, occasionally, lid lag and stare. Laboratory findings are unique because amioda-rone inhibits the conversion of T$_4$ to T$_3$. Thus, even in the euthy-roid patient taking amiodarone, total T$_4$ may be elevated while TSH is normal. In the hyperthyroid patient, the FT$_4$ is markedly elevated and TSH is less than 0.01 mU/L (below the detection limit of the assay). Radioiodine uptake in the iodide-loaded patient is low. It has been difficult to distinguish thyrotoxicosis due to follicular cell hyperfunction (type 1) from that due to fol-licular cell destruction (type 2). Thyroid ultrasound with color Doppler studies may show increased circulation with hyperfunc-tion and decreased blood flow with thyroiditis. Also, the cytokine interleukin (IL)-6 is low in patients with hyperfunction, whereas

TABLE 24–5 Management of amiodarone-induced hyperthyroidism.

(1) Stop amiodarone if possible
(2) Institute beta-blocker therapy if possible
(3) Antithyroid drugs: Methimazole, 40-60 mg/d
(4) Potassium perchlorate, 200 mg every 6 h
(5) Cholestyramine or colestipol, 20-30 g/d
(6) Prednisone, 40 mg/d, for acute thyroiditis (consider monitoring IL-6 levels)
(7) Thyroidectomy

it is markedly elevated in patients with thyroiditis. A recent study suggests that sestamibi scan of the thyroid may be helpful in separating type 1 versus type 2 disease.

Management

Management of amiodarone-induced hyperthyroidism is difficult (Table 24–5). Ideally, amiodarone should be discontinued, but often it cannot be stopped because of the underlying heart disease, and even if it is discontinued, the iodine load persists for several months. Further synthesis of T_4 should be blocked with methimazole in an initial dosage of 40 to 80 mg/d or propylthiouracil in doses of 400 to 800 mg/d followed by a taper to a maintenance dose. β-Adrenergic blockade, if needed (amiodarone, itself, has some β-blocking activity), should be instituted with propranolol or a comparable drug if cardiac status permits it. Potassium perchlorate, which is no longer available in the United States, in a dosage of 250 mg every 6 hours blocks further iodine uptake and lowers intrathyroidal iodide content. Aplastic anemia has occurred in patients on high-dose or long-term potassium perchlorate therapy, so that use of this medication has usually been limited to 1 month. Iopanoic acid (0.5 g orally twice daily) can be used to decrease T_4 to T_3 conversion. Cholestyramine or colestipol in a dosage of 20 to 30 g/d binds T_4 and T_3 in the gut and brings blood levels down more quickly. If there is reason to suspect thyroiditis (elevated serum IL-6 or decreased blood flow on ultrasound), corticosteroid therapy often yields dramatic results. Prednisone is given in a dosage of 40 mg/d for 1 month, gradually tapering the dose over the following 2 months. In those cases where a mixed etiology is suspected (ie, hyperfunction plus thyroiditis), a combination of steroids, antithyroid drugs, and beta-blockers can be used. If medical therapy is unable to control the disease, thyroidectomy results in a permanent cure and may be used as a last resort. Iopanoic acid (discussed earlier) can be used to prepare the patient for surgery. Perioperative mortality may be as high as 8% to 9% due to the underlying cardiac disease.

ACUTE ADRENAL INSUFFICIENCY

Clinical Setting

Acute adrenal insufficiency usually occurs as an acute illness in a patient with chronic adrenal insufficiency (see Chapter 9). The chronic adrenal insufficiency may be primary, due to

destruction of the adrenal glands associated with autoimmune adrenalitis, adrenal leukodystrophy or, rarely, tuberculosis, fungus, or metastatic malignancy. Chronic secondary adrenal insufficiency can be seen in the setting of pituitary or hypothalamic disease. Acute adrenal insufficiency may also occur with bilateral adrenal hemorrhage in a previously healthy individual during the course of septicemia with disseminated intravascular coagulopathy, in a patient with antiphospholipid antibody syndrome, or in a patient receiving anticoagulant therapy. In the patient with known adrenal insufficiency, an acute crisis may be precipitated by inadvertent omission of steroid medication or by the concurrent development of a precipitating illness such as severe infection, acute myocardial infarction, cerebrovascular hemorrhage or infarction, surgery without adrenal support overly strenuous exercise, or severe acute trauma. Acute adrenal insufficiency may also be precipitated by the sudden withdrawal of steroids in a patient previously on long-term pharmacologic steroid therapy with associated adrenal atrophy (ie, secondary adrenal insufficiency). Finally, administration of drugs impairing adrenal hormone synthesis such as ketoconazole, aminoglutethimide, etomidate, or mitotane—or drugs increasing steroid metabolism such as phenytoin or rifampin—may precipitate an adrenal crisis.

The patient presents with an acute onset of nausea, vomiting, hyperpyrexia, abdominal pain, dehydration, hypotension, and shock. A clue to the diagnosis of primary adrenal insufficiency is the presence of pigmentation in unexposed areas of the skin, particularly in the creases of the palms and in the buccal mucosa. The differential diagnosis includes consideration of other causes of cardiovascular collapse, sepsis, and intra-abdominal abscess. Failure of the hypotension to respond to pressors is suggestive of adrenal insufficiency and is an indication for a trial of glucocorticoid therapy.

Diagnosis

Primary adrenal insufficiency is characterized by hyponatremia and hyperkalemia. However, in situations of adrenal crisis, the hyponatremia may be obscured by dehydration. Random serum cortisol determinations are not helpful unless the levels are very low (<5 µg/dL [138 nmol/L]) during a period of great stress. The key diagnostic test is failure of serum cortisol to rise above 20 µg/dL (552 nmol/L) 30 minutes after intravenous injection of 250 µg synthetic ACTH (cosyntropin) (see Chapter 9 for details). This test functions best in the diagnosis of primary versus secondary adrenal insufficiency. At a specificity of 95%, sensitivities are 97% and 57%, respectively (see Chapter 3). Interpretation of the test, particularly in the evaluation of secondary adrenal insufficiency, is complicated in the presence of hypoalbuminemia, a marker for reduced protein-bound cortisol in plasma. In this setting, serum-free cortisol levels provide a more accurate assessment of adrenal function. Random serum free cortisol of greater than 1.8 µg/dL and a Cortrosyn-stimulated value of greater than 3.1 µg/dL in critically ill patients are considered normal but these cutoffs are likely to be assay dependent. There has been increased interest in using a more

physiologic dose of ACTH to perform this stimulation test (1 μg vs 250 μg); however, a recent comparison suggests no particular advantage in the performance characteristics of the low-dose test. Basal serum ACTH is elevated (>52 pg/mL [>11 pmol/L]) in patients with primary adrenal insufficiency but not in patients with secondary adrenal insufficiency due to pituitary or hypothalamic disease. Computed tomography (CT) or sonography of the abdomen reveals adrenal enlargement in patients with adrenal hemorrhage, active tuberculosis, or metastatic malignancy. Atrophy of the adrenals is associated with chronic adrenal insufficiency.

Management

The management of adrenal crisis is outlined in Table 24–6. Hydrocortisone should be administered in a dosage of 100 mg intravenously followed by 50 to 75 mg every 6 hours thereafter. Fluids and Na$^+$ should be replaced with several liters of 5% glucose in normal saline. After the first 24 hours, the dose of intravenous hydrocortisone can be slowly reduced, but intravenous doses should be given at least every 6 hours because of the short half-life (1 hour) of hydrocortisone in the circulation. When the patient can tolerate oral feedings, hydrocortisone can be given orally, but the first oral dose should overlap the last intravenous dose. Alternatively, hydrocortisone can be administered as a continuous infusion at the rate of 10 mg/h for the first 24 hours, followed by a gradual decrease in the dose. Mineralocorticoid is not necessary during the acute replacement period since enough NaCl and glucocorticoid are being administered to treat the mineralocorticoid deficiency. However, in patients with chronic primary adrenal insufficiency, mineralocorticoid supplementation is necessary when shifting to an oral maintenance program (see Chapter 9). After steroid therapy has been instituted, it is extremely important to evaluate and treat the illness that may have precipitated the acute crisis (eg, infection, myocardial infarction).

Prevention of acute adrenal insufficiency in patients with chronic adrenal insufficiency exposed to mild stress can be addressed merely by doubling the daily steroid dose until the condition resolves followed by rapid titration back to prestress doses. More severe stress (eg, severe infection) requires intravenous hydrocortisone in dosages as outlined above or administration of dexamethasone sodium phosphate, 4 mg intramuscularly every 24 hours for two doses. Dexamethasone would replace glucocorticoids but not

TABLE 24–6 Management of adrenal crisis.

(1) Hydrocortisone sodium phosphate or sodium succinate, 100 mg IV stat and then 50-75 mg IV every 6 h for 24 h. Then taper slowly downward over the next 72 h, giving the drug every 4-6 h IV. When patient is tolerating oral feedings, shift to oral replacement therapy, overlapping the first oral and last intravenous doses.
(2) Replace salt and fluid losses with several liters of 5% glucose in normal saline IV.
(3) Patients with primary adrenal insufficiency may require mineralocorticoid (fludrocortisone) when shifted to oral hydrocortisone maintenance therapy.
(4) Diagnose and treat the illness that precipitated the acute crisis.

mineralocorticoid and would not be adequate in the presence of severe dehydration.

PITUITARY APOPLEXY

Pituitary apoplexy is a rare syndrome of violent headache, visual and cranial nerve disturbances, and mental confusion, resulting from hemorrhage or infarction of pituitary tissue.

Clinical Setting

Pituitary apoplexy usually occurs as a sudden crisis in a patient with a known or, more rarely, previously unrecognized pituitary tumor. However, it may occur in a normal gland during or after parturition. Risk factors include cardiac surgery, dynamic pituitary function testing, head trauma, bleeding disorders, hypertension, or anticoagulation therapy. Apoplexy develops in pituitary adenomas due to bleeding and necrosis as they outgrow their blood supply or through compression of infundibular or superior pituitary vessels against the diaphragma sella. The patient presents with severe headache and visual disturbances, often a bitemporal hemianopia due to compression of the optic chiasm. There may be visual and oculomotor defects, either bilateral or unilateral, due to extension of the hemorrhage into the cavernous sinus(es) where cranial nerves II, III, IV, and VI are located. Often there are meningeal symptoms with stiff neck and mental confusion, so that the differential diagnosis includes subarachnoid hemorrhage and meningitis. Finally, there may be symptoms of acute secondary adrenal insufficiency with nausea, vomiting, hypotension, and collapse. A more subacute/chronic presentation (ie, subclinical hemorrhage followed by partial resolution) may be dominated by symptoms of hypopituitarism.

Diagnosis

The diagnosis of pituitary apoplexy is best approached with computed tomography (CT) and magnetic resonance imaging (MRI) of the cranium with multiple sections through the sella turcica. Pituitary enlargement and signs of hemorrhage are diagnostic. CT is most useful in the acute setting (24-48 hours) while MRI is optimal for identifying blood in the subacute setting (4-30 days). In the acute setting, hormonal studies are of academic interest only since therapy should include glucocorticoid support regardless of the acute findings. After appropriate acute management, evaluation of anterior and posterior pituitary function is indicated to evaluate the possibility of permanent hypopituitarism.

Management

Management involves both hormonal and neurosurgical therapy. High-dose dexamethasone, 4 mg twice daily, provides both glucocorticoid support and relief of cerebral edema. Alternatively, hydrocortisone 50 mg IV every 6 hours can be used. Transsphenoidal pituitary decompression often provides dramatic relief of visual and extraocular motor dysfunction and in the level of consciousness. Pituitary hormone secretion also may improve

particularly in those with normal or elevated prolactin levels. The need for emergent surgery is controversial and should be approached on a case-by-case basis. Conservative medical management in less severely affected patients has been shown to result in recovery of both neurological and endocrine function. After the acute episode has subsided, the patient must be evaluated for the possibility of multiple pituitary hormonal deficiencies (see Chapter 4).

DIABETIC KETOACIDOSIS

Clinical Setting

Diabetic ketoacidosis (DKA) occurs in a setting of absolute or relative insulin deficiency. Estimated annual incidence is 4 to 8 episodes per 1000 patients with diabetes. It is seen primarily in patients with type 1 diabetes; however, it is increasingly seen in patients with type 2 diabetes, with approximately one-third of DKA admissions in the USA occuring in type 2 patients. Specific clinical settings should generate a high index of suspicion for the disorder. Interruptions of normal insulin delivery due to purposeful reduction in insulin dosage or interference with the delivery system (eg, kinking in pump tubing) are frequent precipitating events, as are reduced insulin sensitivity in the setting of systemic infection, myocardial infarction, burns, trauma, or pregnancy. In a significant percentage of patients, DKA is the presenting feature of diabetes. In these instances, clinical suspicion and accurate interpretation of the initial laboratory studies will usually lead to the correct diagnosis. Measurement of HbA_{1C} (hemoglobin A_{1C}) levels may help in assessing the chronicity of the diabetes. Mortality in DKA is less than 5% in experienced centers. Prognosis worsens at the extremes of age and in the presence of coma or hypotension.

Diagnosis

DKA is characterized metabolically by two prominent features: hyperglycemia and ketoacidosis (see Chapter 17). Patients with DKA present with evidence of volume contraction (eg, dry mucous membranes, thirst, orthostatic hypotension) and labored breathing (Kussmaul respiration) related to the underlying acidosis. The breath often has a fruity odor, reflecting the presence of acetone. Patients may have abdominal pain mimicking an acute abdomen, nausea, and vomiting. The latter symptoms may be related to elevated gastrointestinal prostaglandins that accrue in the presence of insulin deficiency. Presentation may be dominated by symptoms of the precipitating illness (eg, urinary tract infection, pneumonia, or myocardial infarction).

Plasma glucose levels are elevated, usually to over 250 mg/dL. This reflects impairment in glucose utilization (discussed earlier), increased gluconeogenesis and glycogenolysis, and reduced renal clearance of glucose in the setting of decreased glomerular filtration rate (GFR). Osmotic diuresis related to glucose excretion results in reduction in intravascular volume and depletion of total body water, sodium, potassium, phosphate, and magnesium. In general, the relative depletion of water is roughly twice that of the solutes it contains. Hypertonicity in the extracellular fluid compartment, although typically not as severe as that seen in hyperosmotic nonketotic coma (discussed later), can be significant. Calculated plasma osmolalities greater than 340 mOsm/kg are associated with coma. Older studies suggest that plasma osmolality correlates most closely with state of consciousness in DKA (Figure 24–3); however, more recent studies have forged a stronger link between acidosis and altered consciousness. At present the etiology of coma is unsettled.

Arterial blood pH is low and, in the absence of coexistent respiratory disease, is partially compensated by a reduction in PCO_2. The acidosis is metabolic in origin and accompanied by an anion gap that is calculated by subtracting the combined concentrations of chloride and bicarbonate from serum sodium concentration. Anion gaps greater than 12 mEq/L are considered abnormal. Keto acids account for most of the unmeasured anions that generate the abnormal gap, although under conditions of extreme volume contraction and hypoperfusion, lactate accumulation may also contribute. Levels of serum and urinary ketones (measured using the nitroprusside reagent) are typically high in DKA. It should be recalled, however, that this reagent reacts strongly only with acetoacetate, less strongly with acetone (which is not a keto acid and does not contribute to the anion gap), and not at all with β-hydroxybutyrate. Thus, paradoxically, the most extreme levels of ketoacidosis may be accompanied by relatively modest levels of ketones measured by this method. As a corollary of this, resolution of severe DKA may be linked to transient increases in measurable ketone levels as β-hydroxybutyrate is converted to the more readily detectable acetoacetate. Direct measurement of β-hydroxybutyrate, which is available in most clinical labs, provides a more accurate assessment of ketogenesis.

Serum sodium levels may be high, normal, or low, but in all instances total body sodium is depressed. Estimates of depletion range from 7 to 10 mEq/kg body weight. As blood glucose levels rise in DKA, they create an osmotic gradient that draws water, as well as intracellular solutes, into the extracellular space. This results in moderate hyponatremia, which can be corrected to account for the dilutional effect of the transmembrane flux of water by adding 1.6 mEq/L (more recent studies suggest that the correction factor is closer to 2.4 mEq/L) to the sodium concentration for every 100 mg/dL increment in plasma glucose above a basal concentration of 100 mg/dL.

$$\text{Corrected Na}^+ = \text{Measured Na}^+ + 1.6\left(\frac{\text{Plasma glucose} - 100}{100}\right)$$

The decrease in serum sodium partially offsets the increase in tonicity that accompanies the elevation in plasma glucose. This results in a net increase in plasma osmolality of 2 mOsm/kg H_2O per 100 mg/dL elevation in plasma glucose.

Total body potassium levels are also severely depleted in DKA to an average of 5 to 7 mEq/kg body weight. This results from a number of factors, including exchange of intracellular potassium for extracellular hydrogen ion, impaired movement of K^+ into cells in the insulinopenic state, increased urinary potassium excretion secondary to the osmotic diuresis and, in those instances where
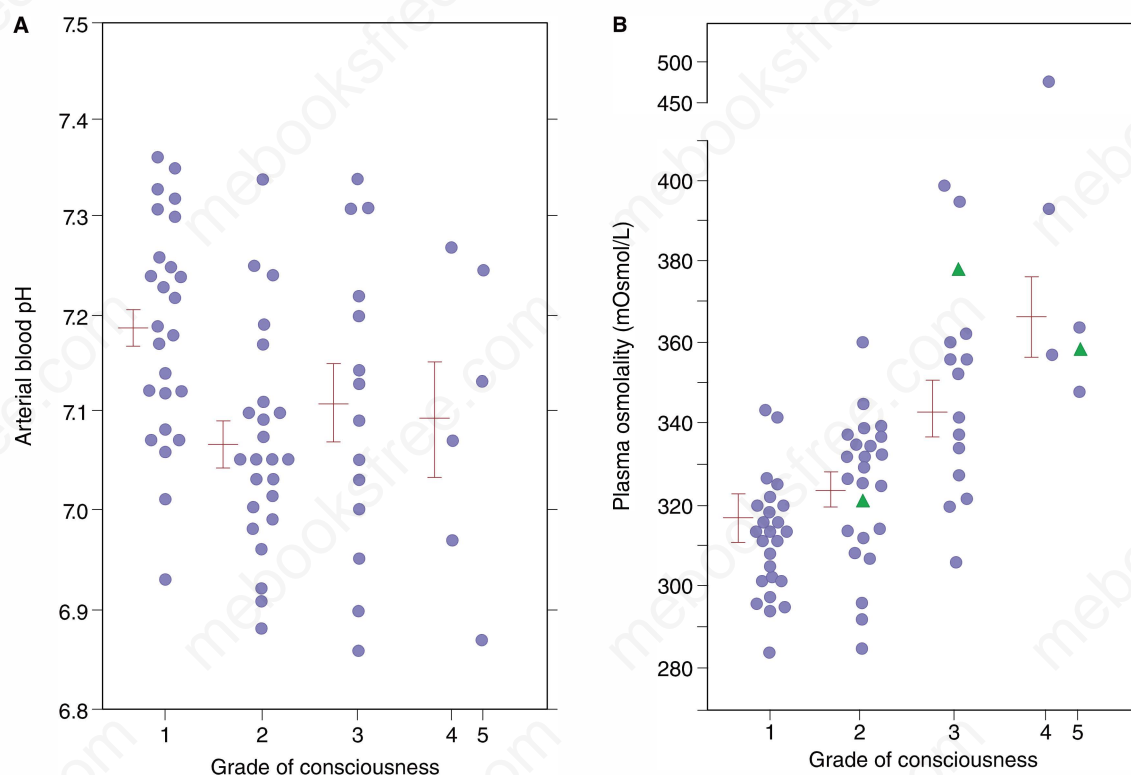
FIGURE 24–3   A: Relationship between state of consciousness and blood pH in patients with diabetic ketoacidosis. **B:** Relationship between state of consciousness and plasma osmolality in diabetic ketoacidosis. (Reproduced with permission from Fulop M, Tannenbaum H, Dreyer N. Ketotic hyperosmolar coma. *Lancet.* 1973 Sep 22;2(7830):635-639.)

intravascular volume contraction is present, secondary hyperaldosteronism. Serum potassium levels may be high, normal, or low depending on the severity and duration of DKA, the status of extracellular fluid volume, and the adequacy of renal perfusion and excretory function. A low-serum potassium at presentation generally indicates severe potassium deficiency and, in the presence of adequate renal function, is an indication for early and aggressive repletion (discussed later). Potassium depletion can result in muscle weakness and cardiac arrhythmias, including ventricular fibrillation.

The H$^+$ excess in DKA titrates endogenous buffer systems including serum bicarbonate, resulting in reduction in concentrations of the latter. Chloride levels may also be low, reflecting the osmotic diuresis alluded to above. Ketone bodies have been estimated to account for one-third to one-half of the osmotic diuresis seen in DKA. Electrolyte depletion is further aggravated by the obligate cation (eg, sodium) excretion required to maintain electrical neutrality. In patients who have maintained adequate hydration during the development of DKA or in those who are aggressively resuscitated with normal saline, chloride levels may be elevated and the anion gap narrowed. This reflects the enhanced clearance of the keto anions in the kidney, converting the system from an anion gap acidosis to a hyperchloremic, nongap acidosis (ie, the hydrogen ion excess persists despite clearance of the anion). Because the excreted keto anions represent a lost source of bicarbonate regeneration, correction of the hyperchloremic

acidosis may proceed slowly. The hyperchloremic acidosis is inconsequential from a clinical standpoint.

Total body magnesium and phosphate levels are also depleted by the osmotic diuresis in DKA. Phosphate depletion is amplified by diffusion of the anion from the intracellular to the extracellular compartment in the absence of insulin. Phosphate depletion can result in muscle weakness, rhabdomyolysis, hemolytic anemia, respiratory distress, and altered tissue oxygenation (due to reduction in 2,3-diphosphoglycerate levels in the red blood cell).

Management

Treatment of DKA is focused on two major objectives. The first is restoration of normal tonicity, intravascular volume, and solute homeostasis. The second is correction of the insulinopenic state with suppression of counterregulatory hormone secretion, glucose production, and ketogenesis and improved utilization of glucose in target tissues. The steps outlined in Table 24–7 provide a general approach to the management of this disorder.

Because depletion of intracellular and extracellular fluids may be severe in DKA (typically in the range of 5-10 L), early and aggressive resuscitation with fluids is mandatory. The goal should be to replace the total volume loss within 24-36 hours with half of the replacement given in the first 12 hours. This is usually initiated with administration of 1 to 2 L of isotonic normal saline (0.9% NaCl) over the first hour of therapy. As intravascular

TABLE 24–7   Management of diabetic ketoacidosis.

Fluid administration
(1) 1-2 L of normal saline over the first hour. Repeat if clinically significant volume contraction persists after the first hour.
(2) Change to half-normal saline, 500-1000 mL/h, depending on volume status. Continue for about 4 h. Decrease rate to 250 mL/h as intravascular volume returns to normal.
(3) Convert fluids to D_5W when plasma glucose falls to 250 mg/dL.

Insulin
(1) Administer 10-20 U of regular insulin IV.
(2) Mix 50 U of regular insulin in 500 mL of normal saline (1 U/10 mL). Discard first 50 mL of infusion to accommodate insulin binding to tubing. Administer through *piggyback* line along with parenteral fluids at a rate of 0.1 U/kg/h.
(3) Double the infusion rate after 2 h if there is no improvement in plasma glucose levels.

Potassium
(1) Administer supplemental potassium chloride once renal function is established; provide 20 mEq/L of fluids for patients who are initially normokalemic, 40 mEq/L for those who are hypokalemic at presentation. In the latter case, hold insulin until serum potassium levels begin to increase.
(2) Gauge subsequent replacement based on serum K^+ measurements at 2-h intervals.

Bicarbonate
(1) Sodium bicarbonate only for patients with blood pH <7.0.
(2) Add 1 ampule of sodium bicarbonate (44 mEq) to 500 mL of D_5W or half-normal saline. Administer over 1 h.

volume is restored, renal perfusion increases, with a consequent increase in renal clearance of glucose and a fall in plasma glucose levels. If volume contraction is severe, another liter of normal saline can be administered. If not, half-normal saline (0.45% NaCl) can be initiated at a rate of 250 to 500 mL/h depending on intravascular volume status. Because water is typically lost in excess of solute in DKA, half-normal saline addresses both the volume depletion and the hypertonicity. Half-normal saline can be continued until intravascular volume has been restored or plasma glucose levels fall to 250 mg/dL, at which point D_5W should be started. The latter maneuver reduces the likelihood of insulin-induced hypoglycemia and avoids the theoretical complication of cerebral edema due to osmotically induced fluid shifts from plasma into the central nervous system. This complication is, in fact, seen rarely in adults and uncommonly in children with DKA. It is important to account for ongoing urinary volume and electrolyte losses in assessing fluid requirements.

Once fluid resuscitation has been initiated, insulin should be administered. Only short-acting insulin should be used. A number of different insulin regimens have demonstrated efficacy in the treatment of DKA; however, a commonly used regimen includes a loading dose (10-20 U) of short-acting insulin intravenously followed by a continuous infusion at a rate of 0.1 U/kg/h. The need for a loading dose is controversial and may not be required in the majority of cases. It is not recommended for children with ketoacidosis. If intravenous access is problematic, maintenance insulin can be given intramuscularly (0.1 U/kg/h). This regimen provides plasma insulin levels in a physiologic range (100-150 mU/mL) with minimal risk of hypoglycemia or

hypokalemia. It restores plasma glucose levels at rates equivalent to those obtained with regimens using higher insulin doses. Plasma glucose levels should fall at a rate of 50 to 100 mg/dL/h. Failure to achieve this end point over a 2-hour period should lead to doubling of the infusion rate with reevaluation an hour later. When plasma glucose concentrations reach 250 mg/dL, the insulin infusion rate should be reduced to 0.02-0.05 units/kg/h and fluid repletion with D_5W is begun to prevent hypoglycemia (discussed earlier). The insulin infusion is continued to suppress ketogenesis, close the anion gap and allow restoration of normal acid-base balance.

As noted earlier, total body potassium stores are depleted in DKA (~3-4 mEq/kg), and plasma K^+ levels fall with treatment. Repletion of K^+ is almost always indicated in management of DKA (one notable exception being DKA that occurs in the setting of chronic renal insufficiency); however, the timing of repletion varies as a function of the plasma K^+ level. If the initial K^+ level is less than 4 mEq/L, K^+ depletion is severe, and repletion should begin with the first administration of parenteral fluids if renal function is adequate. Twenty milliequivalents of potassium chloride can be added to the first liter of normal saline if the serum K^+ is in the 3.5 to 4 mEq/L range; 40 mEq should be added for K^+ levels less than 3.5 mEq/L. Particular attention should be devoted to patients in this latter state, because K^+ levels may plummet to very low levels with initiation of insulin therapy. To avoid this, insulin therapy should be postponed in this group until K^+ repletion has begun and serum K^+ levels are on the rise. The general goal of therapy should be to keep the K^+ in a near-normal range. This may require several hundred milliequivalents of potassium chloride administered over several days.

The administration of bicarbonate in the setting of DKA has been controversial. Acidosis, in addition to increasing ventilatory work (Kussmaul respiration), may also suppress cardiac contractile function. Therefore, restoration of normal pH would seem to make sense in DKA. However, there is considerable risk associated with the use of sodium bicarbonate in this setting, including paradoxical acidification of the central nervous system due to the selective diffusion of CO_2 versus HCO_3^- across the blood–brain barrier and an increase in intracellular acidosis, which may worsen rather than ameliorate cellular function. Volume overload related to the high tonicity (44.6-50 mEq/50 mL) of the bicarbonate solution, hypokalemia resulting from overly rapid correction of the acidosis, hypernatremia, and rebound alkalosis are also potential complications of bicarbonate therapy. In general, pH of 7.0 or greater is not life-threatening to the average patient with DKA and will resolve with appropriate volume expansion and insulin therapy. For pH less than 7.0, many clinicians would argue for a limited administration of sodium bicarbonate. If bicarbonate is used, careful patient monitoring looking for alterations in mental status or cardiac decompensation is indicated. The goal of therapy should be to maintain pH greater than 7.0, not to return pH to normal.

Similarly, phosphate administration, once considered a key component in the management of DKA (estimated deficit ~5-7 mmol/kg), has come under closer scrutiny. Phosphate depletion definitely occurs in DKA for the reasons outlined earlier, and in

the past, repletion of phosphate (much of it as potassium phosphate salts) had been advocated to forestall the development of muscle weakness and hemolysis and to promote tissue oxygenation through generation of 2,3-diphosphoglycerate in erythrocytes. However, the administration of phosphate salts has been associated with the development of clinically significant hypocalcemia and deposition of calcium phosphate precipitates in soft tissues, including the vasculature. Thus, in general, parenteral phosphate repletion is not routinely provided for patients with DKA unless plasma phosphate falls to very low levels (<1 mmol/L). In this case, 2 mL of a mixture of KH_2PO_4 and K_2HPO_4 solution, containing 3 mmol of elemental phosphorus and 4 mEq of potassium, may be added to 1 L of fluids and introduced over 6 to 8 hours. In no instance should all K^+ repletion be in the form of potassium phosphate salts. In general, renewal of food ingestion and insulin therapy complete restoration of total body phosphate stores and return plasma phosphate levels to normal over a period of several days.

In pediatric patients (age <20 years), the need for volume expansion needs to be weighed against the potential risk of cerebral edema secondary to aggressive fluid administration, although this is controversial (discussed later). A recent recommendation suggests 10 to 20 mL/kg/h of normal saline during the first 1 to 2 hours with limitation of total fluid administration to 50 mL/kg over the first 4 hours. The remaining fluid deficit is corrected over the next 48 hours. Normal saline or half-normal saline (depending on serum Na^+ levels) usually accomplishes this with a rate of 5 mL/kg/h. The decrease in serum osmolality should not exceed 3 mOsm/kg H_2O/h. An insulin bolus prior to initiating the insulin infusion (0.1 U/kg/h) is usually not required in children.

Finally, it is necessary to actively seek out and treat the precipitants of DKA when they are identified. This includes appropriate cultures of urine and blood (and cerebrospinal fluid, if indicated) and empiric antibiotic therapy directed against the most likely pathogenic organisms (pending the results of the cultures). The presence of fever is typically a good marker for infection or other inflammatory process because it is not a feature of DKA per se. Elevated white blood cell counts, on the other hand, are frequently seen with DKA alone. Hyperamylasemia is common but rarely reflects pancreatitis—the amylase is usually of salivary origin. Other precipitants should also be sought. Myocardial infarction, which is often clinically silent in diabetic patients, is an uncommon but life-threatening precipitant of DKA in patients with established diabetes.

Complications

Aggressive resuscitation with isotonic or hypotonic fluids is a theoretical but uncommon cause of fluid overload during management of DKA. Careful attention to the cardiovascular examination, chest x-ray, and urine output should aid in preventing this complication.

Hypoglycemia is relatively rare in the current era given the low doses of insulin used in management and appropriate initiation of glucose-containing fluids as plasma glucose levels fall below 250 mg/dL.

Cerebral edema due to rapid correction of plasma hypertonicity has usually been reported with plasma glucose levels below 250 mg/dL. Clinically significant cerebral edema is relatively uncommon in adult patients. Milder forms of cerebral edema have been noted in many patients being treated for DKA but have not been strongly correlated with changes in extracellular tonicity. At present, in a symptomatic adult patient, it would appear prudent to treat hypertonicity exceeding 340 mOsm/kg aggressively with hypotonic fluids to avoid complications related to plasma hyperviscosity. Further correction from that point to normal plasma osmolality (about 285 mOsm/kg) can probably be accomplished in slower fashion over several days. Cerebral edema occurs in 1% to 2% of children with DKA, frequently with devastating results. Approximately one-third of children with clinically significant cerebral edema die during the acute illness, and another third sustain permanent neurologic impairment. The predilection for young children may reflect, in part, immaturity of the autoregulatory mechanism that governs cerebral blood flow. There is increased risk in children less than 5 years of age, and those with low Pco_2 or high blood urea nitrogen at presentation. Cerebral edema in children may be associated with high initial rates of fluid resuscitation (>4 L/m^2/d) and rapid falls in plasma sodium (or corrected sodium) concentration, although it can occur in clinical settings without an apparent cause, and mild degrees of cerebral edema have been noted in DKA even prior to initiation of therapy. In the absence of definitive trial data to guide therapy, lower rates of fluid administration (<2.5 L/m^2/d) with volume resuscitation spread over a longer time interval would seem appropriate if the clinical situation permits. Insulin administration has also been linked to the development of cerebral edema. Current recommendations do not include the use of insulin boluses in children with DKA. When signs of cerebral edema appear—deterioration in level of consciousness, focal neurologic signs, hypotension or bradycardia, sudden decline in urine output after an initial period of apparent recovery following treatment for DKA—fluid administration should be reduced and mannitol (0.2-1 g/kg given intravenously over 30 minutes) should be administered with repetition at hourly intervals based on response. CT or MRI scan of the brain can be done once therapy has been initiated to confirm the diagnosis. Hyperventilation has not been shown to alter the course of this complication once it develops.

Patients with DKA are also prone to develop acute respiratory distress syndrome, presumably reflecting the sequelae of a damaged pulmonary endothelium and elevated capillary hydrostatic pressures following fluid resuscitation. Patients who present with rales at the time of initial diagnosis may be at higher risk for the development of this complication. Patients may also be at increased risk for development of pancreatitis as well as systemic infection, including fungal infections (eg, mucormycosis).

Abdominal pain and gastric stasis seen in DKA may put a semi-stuporous patient at risk for aspiration. Up to 25% of patients with DKA have emesis that may be guaiac positive. The latter finding appears to result from hemorrhagic gastritis. Patients who are believed to be at risk with regard to airway protection should have a nasogastric tube in place for evacuation of stomach contents.

Finally, patients with DKA are at risk for recurrence of the disorder if insulin is withdrawn prematurely. The current infusion protocols, because they raise plasma insulin only to physiologic levels, have a very short half-life for control of blood glucose and ketogenesis. Premature cessation of insulin therapy before depot insulin (eg, NPH or glargine insulin) can exert its effect may allow the patient to regress into ketoacidosis. To preclude this possibility, subcutaneous regular and intermediate-acting insulin should be provided on the morning when feeding is to be resumed. The insulin drip should be continued for 1 hour following this injection to provide coverage until the depot insulin becomes effective.

HYPEROSMOLAR NONKETOTIC COMA

Clinical Setting

Hyperosmolar nonketotic coma, like DKA, is a consequence of uncontrolled diabetes mellitus; however, a number of features of this disorder clearly distinguish it from DKA. First, plasma glucose levels in ketoacidosis are usually in the 250 to 400 mg/dL rather than the 700 to 1000 mg/dL range that can be seen with hyperosmolar nonketotic coma. Second, the time course for development of hyperosmolar nonketotic coma typically takes a week or longer, whereas ketoacidosis can develop over 1 to 2 days. Third, ketosis is rare in hyperosmolar nonketotic coma. Circulating insulin levels are probably adequate to control ketogenesis in the latter, although they are incapable of establishing euglycemia. Fourth, hyperosmolar nonketotic coma tends to occur more commonly in the elderly population, often in those receiving chronic care who have difficulty reporting symptoms or maintaining adequate hydration. Mortality remains high in hyperosmolar nonketotic coma (perhaps as high as 20% vs 1%-2% in DKA); in both cases, mortality increases with age.

Precipitants of hyperosmolar nonketotic coma include many of the same complicating illnesses that lead to DKA. Infections (pneumonias are said to be the most common precipitating infection in 40%-60% of cases, with urinary tract infections representing 5%-16% of the total), myocardial infarction, cerebrovascular accident, pancreatitis, burns, heat stroke, and endocrine dysfunction (eg, Cushing syndrome, acromegaly) are frequently associated with hyperosmolar nonketotic coma. Up to 20% of patients with hyperosmolar nonketotic coma do not have a previous diagnosis of diabetes mellitus. Administration of hyperosmolar fluids (eg, tube feedings, total parenteral nutrition, peritoneal dialysis) can precipitate hyperosmolar nonketotic coma, as can medications that impair insulin secretion or action (eg, β-adrenergic blocking agents, phenytoin, corticosteroids, diazoxide, or atypical antipsychotics). Diuretics, particularly thiazides, can reduce intravascular volume, reduce GFR (the dominant mechanism for glucose clearance), and activate counterregulatory hormones (eg, catecholamines), all of which promote development of the hyperosmolar state. Finally, limited access to free water intake, particularly in patients who are dependent on others to provide access to water, is a major determinant in the progression of the hyperosmolar state.

These factors interact in a highly variable fashion to promote hyperosmolar nonketotic coma. In a typical scenario, poorly controlled diabetes—or undiagnosed diabetes presenting for the first time—is aggravated by coincident infection (or other precipitant). This results in significant hyperglycemia as the balance of counterregulatory hormones versus insulin shifts in favor of the former. Hyperglycemia promotes increased insulin resistance and further elevations in blood glucose levels. This leads to an osmotic diuresis in the kidney as plasma glucose levels exceed the threshold for tubular reabsorption. This threshold typically increases with age, leading to even higher plasma glucose levels in the elderly. The osmotic diuresis results in a progressive loss of water and, to a lesser degree, solutes (eg, Na^+, Cl^-, K^+) in the urine. As with DKA, the water loss is buffered initially by an osmotically driven movement of water out of the cells into the extracellular compartment. As the diuresis continues, however, contraction of intravascular volume ensues and glomerular filtration falls, shutting off the body's primary mechanism for controlling plasma glucose levels in this setting. Glucose levels increase dramatically; often to extraordinarily high levels (normal renal excretory function generally limits elevations in plasma glucose to 500-600 mg/dL). Nausea and vomiting, related to the infection, uremia, or hyperosmolality per se, may further aggravate intravascular volume contraction. Factor into this an elderly patient with age-dependent suppression of the thirst mechanism and difficulty communicating a sense of thirst to caregivers, and the stage is set for the severe dehydration of hyperosmolar nonketotic coma.

Diagnosis

Current diagnostic criteria from the American Diabetes Association include plasma glucose greater than 600 mg/dL, effective plasma osmolality greater than 320 mOsm/kg and absence of ketoacidosis. The diagnosis is usually based on strong clinical suspicion and laboratory assessments of plasma glucose levels and serum osmolality. The typical patient may be a patient with known type 2 diabetes (often taking an oral hypoglycemic agent), who has shown a subtle but steady deterioration over several days immediately preceding admission. The patient demonstrates significant dehydration, somnolence, stupor, or coma. The course may be marked by a precipitating illness, as described earlier. Tachycardia and low-grade fever may be present, but blood pressure and respiratory rate are normal. Urine output typically is reduced to low levels as the hyperosmolar state progresses.

The clinical presentation is usually dominated by the central nervous system findings. Patients with hyperosmolar nonketotic coma are lethargic and weak, with an altered sensorium. True coma is less common and usually does not appear until plasma osmolality is significantly elevated (discussed later). Hallucinations are occasionally present, and seizures, particularly focal seizures, may occur in 25% of patients. Focal findings suggesting cortical ischemia may also be seen. A minority of these reflect true cerebrovascular accidents related to thromboembolic complications (discussed later). Most represent low-flow states in areas of baseline cerebrovascular ischemia that recover with correction of the metabolic abnormality.

Plasma glucose levels are typically high (occasionally >1000 mg/dL). Measured plasma osmolality is often greater than 350 mOsm/kg. These measurements usually include the contribution of urea, which may accumulate to significant levels in the setting of volume contraction and prerenal azotemia; however, urea does not contribute to the osmotic force that drives fluid movement across cellular membranes because it is freely permeable in these membranes and distributes itself in equivalent concentrations in the intracellular and extracellular compartments. A more accurate measure of the effective osmolality, or tonicity, is obtained from the following formula:

$$\text{Effective osmolality} = 2Na^+ + \frac{\text{Glucose}}{18}$$

where Na^+ and glucose are the concentrations of sodium ion and glucose, respectively. Normal effective osmolalities are in the range of 275 to 295 mOsm/kg. Hyperosmolar nonketotic coma is diagnosed with an effective osmolality greater than 320 mOsm/kg, whereas coma itself is seen with osmolalities greater than 340 mOsm/kg. Coma at effective osmolalities less than or equal to 340 mOsm/kg suggests some other cause (eg, meningitis, cerebrovascular accident, or other metabolic abnormality).

Anion gap acidosis is usually not a major component of hyperosmolar nonketotic coma, because ketogenesis is suppressed (discussed earlier). If gap acidosis is present, it may suggest lactate accumulation due to tissue hypoperfusion, low-output state, metformin toxicity, or coexistent renal failure or toxic ingestion. Measurement of serum lactate levels, standard renal function tests, or a screen for toxic substances in blood (if a suggestive history is obtained) should help discriminate among these possibilities.

Management

Therapy is in some ways similar to that for DKA except that there is less concern about correction of the acidosis, which is mild or nonexistent with most conventional cases of hyperosmolar nonketotic coma, and greater emphasis is placed on restoration of intravascular volume and serum osmolality. It is also important to address the potentially life-threatening precipitants of nonketotic coma.

Resuscitation of intravascular volume is initiated with parenteral fluids. Over the first hour 1 to 2 L of normal saline are administered. If the patient is profoundly hyperosmolar (>330 mOsm/L), half-normal saline (0.45% NaCl) should be substituted for normal saline. Volume status is reassessed based on blood pressure, urine output, or central venous pressure. If volume contraction persists, normal saline (or half-normal saline if hypertonicity persists) can be continued at a rate of 1 L/h. Once blood pressure and urine output have been restored, hypotonic fluids (eg, half-normal saline) can be substituted for isotonic saline at a rate of 250 to 500 mL/h depending on the need for continued volume resuscitation. The total free water deficit can be approximated using the following formula:

$$\text{Water deficit} = \frac{\text{Plasma osmolality} - 295}{295} \times 0.6 \,(\text{Body weight})$$

where plasma osmolality is in milliosmoles per kilogram, body weight in kilograms, and water deficit in liters. Fluid rates should be adjusted to correct half of the free water deficit in the first 12 hours and the remainder over the ensuing 24 to 36 hours. As with DKA, glucose-containing fluids (eg, D_5W) should be started when plasma glucose levels fall to 250 mg/dL.

Insulin therapy is of secondary importance in the management of hyperosmolar nonketotic coma. It is imperative that insulin therapy not be initiated until volume resuscitation is well under way (eg, following 1-2 L of saline). Insulin promotes movement of glucose, electrolytes, and water into the intravascular compartment. In the absence of adequate volume resuscitation, this can lead to hypotension and cardiovascular collapse. Therapy should be initiated with a loading dose of 10 to 20 U intravenously followed by a drip delivering 0.1 U/kg/h. Given the theoretical risk of brain edema with rapid reduction in plasma osmolality, it is recommended that glucose levels be maintained at approximately 300 mg/dL in the acute setting. Conversion back to the outpatient regimen once the acute event has resolved can be accomplished using a strategy similar to that outlined earlier for DKA.

Electrolytes (ie, Na^+, K^+, Cl^-, PO_4^{3-}, and Mg^{2+}) are significantly depleted in hyperosmolar nonketotic coma due to the osmotic diuresis. These should be repleted as needed (beginning with the first liter of fluids if necessary). Once again, parenteral phosphate should be administered with care, keeping serum phosphate above 1 mg/dL until feeding can re-establish phosphate balance.

Complications

A. Thromboembolic events Coagulopathies related to increased platelet aggregation, hyperviscosity of circulating blood, or disseminated intravascular coagulation can develop in hyperosmolar nonketotic coma. The most appropriate therapy is resuscitation of intracellular volume and treatment of systemic infections. Focal neurologic findings that fail to improve with fluid resuscitation should be further investigated with appropriate consultation and imaging studies, if indicated. Some investigators have recommended low-dose heparin anticoagulation in patients with hyperosmolar nonketotic coma to guard against thromboembolic sequelae. If such therapy is initiated, patients should be closely monitored for development of gastrointestinal bleeding.

B. Cerebral edema This is a serious—although fortunately uncommon—complication of fluid resuscitation in hyperosmolar nonketotic coma (and DKA; discussed earlier). It occurs more commonly in children than in adults and frequently follows an overly aggressive management strategy involving administration of large amounts of parenteral fluids. The pathogenesis is not completely defined but probably relates to the increase in cortical capillary hydrostatic pressure and the osmotic gradient engendered by the aggressive use of hypotonic fluids in this setting. Cerebral edema rarely appears with serum glucose levels above 250 mg/dL. Appropriate introduction of glucose-containing

fluids into the management strategy when plasma glucose approaches this level is an effective way to guard against this complication.

Clinically, brain edema develops after an initial period of improvement. It is often heralded by the development of headache, altered mental status, and seizure activity. If untreated, this can progress to herniation, with respiratory arrest and death. Recognition of the syndrome and appropriate treatment with mannitol, dexamethasone, and furosemide can be life saving in this setting.

HYPERCALCEMIC CRISIS

Clinical Setting

Severe hypercalcemia, defined arbitrarily as a total Ca^{2+} greater than 14 mg/dL, can occur in the setting of a known hypercalcemic illness but may represent the initial manifestation of such a disorder. Patients present initially with polyuria and polydipsia and, with a more protracted course, develop evidence of intravascular volume contraction with decreased urine output. Alterations in the sensorium dominate the clinical picture and range from behavioral changes and drowsiness to stupor and coma. Bradyarrhythmias and heart block are major cardiac sequelae of hypercalcemia. Hypercalcemia also potentiates digoxin activity, increasing the risk of cardiac glycoside toxicity. Gastrointestinal complaints are prominent. Anorexia, nausea, and vomiting, which further aggravate the volume contraction, are frequently present. Abdominal pain may be of sufficient intensity to mimic an acute abdomen.

Most chronic hypercalcemia is caused by primary hyperparathyroidism and is usually detected as a result of routine laboratory screening. Cancer-related hypercalcemia, most of which is caused by parathyroid hormone–related protein (PTHrP), is a less frequent but important cause of chronic hypercalcemia (see Chapter 21). In acute hypercalcemic crises, malignancy emerges as the major cause of the elevations in serum calcium. Other causes of hypercalcemia (eg, vitamin D intoxication, thiazides, or Addison disease) are either too uncommon or cause such small increments in serum calcium that they rarely need to be considered in the differential diagnosis (see Chapter 8).

Hypercalcemia due to any cause creates a state of nephrogenic diabetes insipidus by uncoupling vasopressin from its receptor-effector system in the kidney (Table 24–8). It also promotes an increase in urinary sodium excretion. This results in a salt and water diuresis that eventually promotes intravascular volume contraction and reduced GFR, effectively suppressing the only route of egress for calcium mobilized from bone or transported across the gut lumen. Thus, volume contraction is in large part responsible for the very high levels of serum calcium found in hypercalcemic crisis. By inference, resuscitation of intravascular volume (discussed later) represents an excellent initial intervention to improve renal perfusion and tubular clearance of Ca^{2+}.

TABLE 24–8 Pathogenesis of hypercalcemic crisis.

Calcium mobilized from bone
↓
Increased urinary calcium
Serum calcium remains normal initially
↓
Nephrogenic diabetes insipidus
+
Increased natriuresis
Decrease in GFR
↓

Decreased clearance of calcium from serum	Continued calcium mobilization from bone
↓	↓

Increased serum calcium
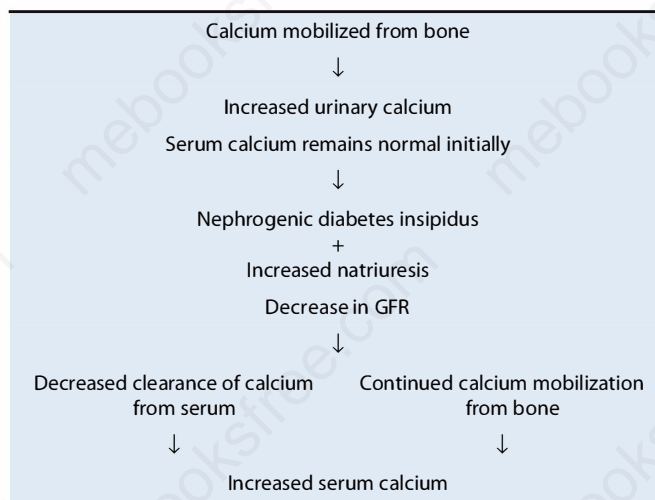
Hypercalcemic crisis should be considered in any patient with disseminated malignancy, particularly squamous carcinomas of the head, neck, and lung. This is especially true when there is a change in the patient's mental status or general clinical condition that cannot be explained by tumor progression, infection, or other metabolic abnormality (eg, uremia). It should also be considered in patients with known primary hyperparathyroidism, particularly in a clinical setting characterized by vomiting, diarrhea, or dehydration (eg, due to thiazide therapy).

Diagnosis

The diagnosis is confirmed by measurement of a total or ionized serum calcium level. The latter may be preferable in the presence of low serum albumin, the predominant Ca^{2+}-binding protein in blood, because hypoalbuminemia may mask an elevation in the free fraction if only total calcium levels are assessed. Albumin-corrected calcium levels can be derived by increasing the total calcium by 0.8 mg/dL for each 1 g/dL decrease in serum albumin, based on a normal albumin level of 4 g/dL. The albumin-adjusted calcium does not always correlate with the measured ionized calcium, however, and should not be relied on for more than a rough estimate of the free calcium fraction, particularly if reliable ionized calcium measurements are available.

On the initial presentation with hypercalcemia, plasma samples should be sent for measurement of intact parathyroid hormone (PTH), PTHrP, and, if the clinical setting is suggestive, 25-hydroxyvitamin D or 1,25-dihydroxyvitamin D levels. Primary hyperparathyroidism is a common disease, and even in a patient with known malignancy it should be excluded as a potentially curable cause of the hypercalcemia.

Management

Intravenous fluids represent the first avenue of approach for management of severe hypercalcemia (Table 24–9). Normalization of intravascular volume improves GFR and increases renal excretion

TABLE 24–9 Therapy for hypercalcemic crisis.

0.9% NaCl (250-500 mL/h) plus furosemide if required
↓
Calcitonin (2-4 mg/kg SC or IM every 8 h) with or without prednisone
↓
Pamidronate (60-90 mg IV over 1 to 2 h)
or
↓
Zoledronic acid (4 mg IV over 15 min)
↓
Evaluate need for retreatment

of calcium. Sodium and calcium handling are closely linked in the distal nephron. Fluid resuscitation with normal saline (0.9% NaCl) both restores GFR and promotes natriuresis and calciuresis by controlling transporter mechanisms responsible for sodium and calcium handling in the distal nephron. Over the first hour 500 to 1000 mL of normal saline is given, with rates of 250 to 500 mL/h thereafter, depending on the state of volume contraction. The latter can be assessed based on clinical examination, urine output, and assessment of renal function. Several liters of fluid are frequently required before intravascular volume is restored. Continued infusions should be matched with urine outputs to avoid fluid overload. Loop diuretics (eg, furosemide) may be used to accomplish this in patients with an underlying predilection toward fluid retention (eg, congestive heart failure). Saline and loop diuretics can increase urinary calcium excretion by as much as 800 mg/d. This is typically accompanied by a moderate but significant reduction in serum calcium levels (1-3 mg/dL). Careful attention should be devoted to detection of signs of fluid overload. Potassium and magnesium depletion related to the diuresis should be corrected. If loop diuretics are used, it is important that the volume of normal saline administered should at least match urine output. Diuretic-induced volume contraction may lead to reduced GFR and worsening hypercalcemia.

At this point, more definitive and specific therapy should be introduced (see Chapter 8). Severe hypercalcemia is almost always a result of increased mobilization of calcium from bone. Therefore, most effective therapies for hypercalcemia have been directed against the osteoclasts of bone. Bisphosphonates represent the mainstay of therapy. Pamidronate administered at a dose of 60 to 90 mg in 250 mL of saline over 1 to 4 hours (4 hours for outpatients) is effective in reducing serum calcium levels, often into the normal range. A newer generation bisphosphonate, zoledronic acid (4 mg intravenous infusion over 15-30 minutes) is heavily used for management of hypercalcemia. In comparison to pamidronate, it is more potent, with a similar safety profile. In addition, it has the advantage of a shorter infusion period and longer duration of antihypercalcemic activity in head-to-head comparisons with pamidronate. The effect of intravenous bisphosphonates can take 2 to 4 days to peak, and the duration of the response is

variable, lasting from 1 week to several months. Retreatment for recurrent hypercalcemia is usually successful. Side-effects of therapy include inflammation at the infusion site; low-grade fever with flu-like syndrome; osteonecrosis of the mandible and maxilla with long-term or frequent infusions; renal insufficiency; and transient depression of serum calcium, phosphate, and magnesium. Bisphosphonates are also attractive because of their efficacy in controlling pain and fracture in myeloma and osteolytic metastases from breast cancer and other solid tumors. The mechanism here is thought to involve suppression of tumor-induced osteoclastic activity in the neighborhood of the metastases.

Denosumab can be an effective alternative to bisphosphonate therapy. It is a monoclonal antibody (against the RANK ligand; see Chapter 8) which acts through interference with the RANK ligand-dependent osteoclastic bone resorption. Denosumab has been shown to be effective in managing hypercalcemia that is refractory to bisphosphonates. Standard dose is 120 mg administered subcutaneously with repeat dosing no sooner than 7 days following the first administration. Reduction in serum calcium is seen within 2 to 4 days following administration of denosumab. Denosumab has been associated with bone pain, nausea, diarrhea, and osteonecrosis of the jaw—though the latter is typically associated with more prolonged use of this agent. There is also a theoretical risk of infection which is based on the role of the RANK ligand in signaling the inflammatory response.

Calcitonin has a long history of use in the management of hypercalcemia. It has a direct effect on the osteoclast to suppress bone resorption. It is given either subcutaneously or intramuscularly at a dose of 2 to 4 U/kg body weight every 6 to 12 hours after administration of a test dose to exclude hypersensitivity to the drug. In general, the response to calcitonin, which should be noted after 6 to 12 hours, is modest in magnitude (decrease in serum calcium levels of 1-2 mg/dL) and declines with increasing duration of therapy (tachyphylaxis). Coadministration of glucocorticoids with calcitonin may limit this latter effect and extend the duration of the hypocalcemic effect. Calcitonin is useful largely as adjunctive therapy in controlling hypercalcemia in the acute setting until the effects of more powerful but slower acting agents (eg, bisphosphonates) become available.

Cinacalcet, a calcium receptor agonist, can be useful in the management of PTH-dependent hypercalcemia such as that seen with parathyroid carcinoma. Doses are typically in the range of 30 to 60 mg bid or higher.

Plicamycin is a tumoricidal antibiotic with pronounced hypocalcemic properties at nontumoricidal doses. It presumably targets the osteoclast and its bone-resorptive activity. It is administered as an infusion at a dosage of 15 to 25 μg/kg body weight over 4 to 24 hours. Calcium levels fall, often into the normal range, within 24 to 48 hours in a majority of patients treated. Plicamycin has significant renal and hepatic toxicity, and it induces platelet abnormalities that may result in clinically significant hemorrhage. Toxicity tends to increase with repeated administration, limiting chronic use of the drug. Given the ready availability of less toxic treatment modalities, plicamycin is rarely used in the management of severe hypercalcemia.

Gallium nitrate has shown efficacy in the management of tumoral hypercalcemia in clinical trials. Like the bisphosphonates and plicamycin, gallium is apparently targeted at the osteoclast. Patients are infused with gallium nitrate at a dosage of 200 mg/m^2 over a 5-day period. Gallium's major limitations are the duration of the infusion (often requiring hospitalization) and the potential for significant nephrotoxicity, particularly if it is used in the setting of other nephrotoxic agents. For these reasons it, like plicamycin, has largely been supplanted by newer agents for management of hypercalcemia. Other nonselective treatment modalities that are available for management of hypercalcemic crisis include steroids, phosphate, and dialysis. Aside from their potentiation of calcitonin's effects, as described earlier, steroids are usually effective only in the management of hypercalcemia due to lymphoproliferative disease (eg, multiple myeloma) or vitamin D toxicity (eg, due to ingestion of vitamin D or disseminated granulomatous disease). They are best suited for chronic management of hypercalcemia associated with these disorders. Parenteral phosphate effectively reduces serum calcium levels, although often at the expense of deposition of calcium phosphate salts in parenchymal tissues, and cannot be recommended. Peritoneal or hemodialysis against a low-calcium bath is also very effective in reducing serum calcium levels and is the treatment of choice for severe hypercalcemia in renal failure patients incapable of tolerating or responding to saline diuresis. In clinically tenuous patients not in renal failure undergoing maximal saline diuresis, persistent hypercalcemia may be an indication for hemodialysis as a tool to bridge the interval until more definitive therapy (eg, bisphosphonates) can achieve its maximal effect.

Finally, some effort should be taken to addressing the primary cause of the hypercalcemia. In some instances (eg, disseminated malignancy), therapeutic options may be limited and ineffective in controlling serum calcium levels. In other instances (eg, primary hyperparathyroidism), a definitive surgical approach can be curative and limit further morbidity. Thus, careful investigation of the source of the hypercalcemia is warranted. Such investigation—even in patients with an obvious potential source of hypercalcemia (eg, malignancy)—may identify correctable problems and, at the very least, assist in the development of long-term management strategies once the acute crisis has resolved.

ACUTE HYPOCALCEMIA

Clinical Setting

Hypocalcemia may be seen in a number of disorders affecting the synthesis or action of PTH or vitamin D or following sequestration of calcium into a functionally inaccessible compartment (see Chapter 8). Many of these represent chronic illnesses where hypocalcemic symptoms develop insidiously or where the complication of hypocalcemia is anticipated early and appropriate treatment initiated prior to acute decompensation. However, in selected situations, acute hypocalcemia may dominate the clinical presentation. Appropriate recognition of the high-risk clinical setting should lead to earlier diagnosis and therapeutic intervention with reduced morbidity and mortality.

Probably the most important cause of PTH-deficient hypocalcemia occurs in the postoperative setting following neck surgery for treatment of malignancy or resection of adenomatous or hyperplastic parathyroid glands. This may reflect accidental or purposeful (eg, radical neck dissection) removal of all functioning parathyroid tissue or inadvertent vascular compromise of tissue left in the neck. Residual normal parathyroid tissue can also be functionally atrophied in a patient undergoing surgery for hyperparathyroidism. It can take 1 to 2 days for PTH secretion and calcium levels to return to normal following resection of the adenoma.

Magnesium deficiency can compromise both PTH secretion from the parathyroid gland and PTH action at target tissues in the periphery. Magnesium repletion in patients with low serum magnesium and calcium should be undertaken before launching an exhaustive workup of the hypocalcemia.

Hypocalcemia can be seen in chronic renal insufficiency. Hyperphosphatemia and reduced 1-α-hydroxylase activity result in deficient 1,25-dihydroxyvitamin D$_3$ generation and consequent hypocalcemia. Aggressive management of the hyperphosphatemia (eg, with phosphate binders) and calcitriol are generally useful in promoting normal calcium balance.

Acute sequestration of calcium into bone or nonphysiologic compartments can lead to severe hypocalcemia. Hypocalcemia following removal of a parathyroid adenoma may reflect a hypoparathyroid or aparathyroid state, as discussed earlier. Alternatively, in patients with severe osteitis fibrosa cystica, the bones, in the absence of PTH-driven bone resorption, serve as a sink for extracellular calcium deposition, as previously unmineralized osteoid becomes calcified. This results in the syndrome termed hungry bones or recalcification tetany. It is usually distinguished from PTH deficiency by measurements of the hormone (PTH is typically elevated in the hungry bones syndrome). A similar sequestration phenomenon is seen in osteoblastic metastases (eg, in breast or prostatic carcinoma). Typically, osteoclastic activity in these metastases is abrogated through some specific therapeutic intervention, leaving unmineralized matrix to calcify at the expense of extracellular calcium levels.

Sequestration may also occur in nonphysiologic settings such as the peritoneal cavity in acute pancreatitis, where deposition of calcium soaps leads to a subsequent reduction in serum calcium levels. A similar phenomenon occurs in damaged muscle following rhabdomyolysis. Deposition of calcium salts in damaged muscle beds leads to a reduction in serum calcium levels. Interestingly, serum calcium returns to normal or even elevated levels during the recovery phase, reflecting dissolution of the precipitates as the muscle undergoes repair.

Hypocalcemia occurs in the setting of acute systemic illness with a prevalence of nearly 80% in the ICU setting. It has been linked to the cytokine-mediated inflammatory response that frequently dominates this clinical setting. Other potential contributors include increased calcium binding to albumin (eg, as seen with alkalosis), relative PTH deficiency, and decreased renal 1-α-hydroxylase activity. It has also been associated with specific drugs, including antineoplastic agents such as doxorubicin and cytarabine and other agents such as ketoconazole, pentamidine, and foscarnet.

Diagnosis

Symptomatic hypocalcemia presents with a predictable constellation of signs and symptoms. Most of these findings are related to the increased neuroexcitability that is seen with reductions of extracellular calcium levels (Table 24–10). Symptoms frequently begin as circumoral paresthesias or paresthesias of the fingers or toes. This is followed by increased muscle cramping and spasm (particularly carpopedal spasm) (see Figure 8–18) and diffuse hyperreflexia. The increased propensity for muscle spasm can provoke generalized tetany which if extended to the laryngeal muscles can lead to laryngospasm and respiratory arrest. Increased excitability in the central nervous system can result in seizures, particularly in patients with a history of seizure disorder.

Physical examination looking for evidence of neuromuscular hyperexcitability is often revealing (see Table 24–10). Chvostek sign is evoked by repetitive tapping of the area overlying the facial nerve approximately 2 cm anterior to the ear lobe below the zygomatic arch. A positive test is contraction of musculature innervated by the facial nerve. The extent of the contraction is roughly proportionate to the severity of the hypocalcemia. Trousseau sign is triggered by inflating a blood pressure cuff on an upper extremity to a level that roughly equates with the systolic blood pressure for 3 to 5 minutes. Spasm of the hand musculature (see carpopedal spasm, discussed earlier) due to transient ischemia of hyperexcitable nerves innervating the hand is regarded as a positive test. Of the two, a positive Trousseau sign is regarded as more specific for hypocalcemia than is the positive Chvostek sign.

Hypocalcemia can also have significant effects on cardiovascular function, including decreased blood pressure, impaired cardiac contractility, and conduction disturbances. On the electrocardiogram, significant hypocalcemia can manifest as prolongation of the QT interval. In extreme cases this can lead to ventricular arrhythmias, including torsades de pointes. Evidence of sequelae of chronic hypocalcemia may also identify individuals who are at risk for acute hypocalcemia. Subcapsular cataracts and basal ganglia calcification are features associated with long-standing hypocalcemia.

Measurement of total serum calcium, albumin, or ionized calcium, if available, leads quickly to the correct diagnosis.

TABLE 24–10   Symptoms and signs of acute hypocalcemia.

Symptoms
Perioral numbness
Tingling paresthesias in distal extremities
Hyperreflexia
Muscle cramps
Carpopedal spasm
Laryngospasm
Seizures
Coma

Signs
Chvostek sign
Trousseau sign
Hypotension
Bradycardia
Prolonged QT interval
Arrhythmias

Measurement of plasma PTH, 25-hydroxyvitamin D, and 1,25-dihydroxyvitmin D assists in identifying deficiency of PTH secretion, adequacy of vitamin D stores, or impaired synthesis of bioactive vitamin D, respectively. Mg^{2+} levels should be obtained to exclude hypomagnesemia. Magnesium depletion can present with the same symptoms as hypocalcemia. If present, magnesium repletion should be initiated before additional diagnostic tests are undertaken.

Examination of the medication list looking for drugs with hypocalcemic properties (eg, bisphosphonates, calcitonin, asparaginase, cisplatin, foscarnet) should be performed.

Measurement of serum amylase and lipase should aid in identifying pancreatitis as a source of the hypocalcemia. Creatine kinase and aldolase levels can be used to exclude the presence of rhabdomyolysis.

One area of potential confusion is the evaluation of patients in the immediate postparathyroidectomy period. Differentiating a hypoparathyroid or aparathyroid state from hungry bones syndrome can be difficult acutely. As pointed out earlier, measurement of plasma PTH levels (high in hungry bones syndrome and low in hypoparathyroidism) provides the most definitive separation of the two, but outside of referral centers this typically takes several days and, therefore, is usually not helpful in the acute setting. Measurement of serum phosphate (elevated in hypoparathyroidism and low-normal or low in hungry bones syndrome) or urinary phosphate (often low or absent in hypoparathyroidism) may be useful in identifying the source of hypocalcemia. Fortunately, the treatment of hypocalcemia in these two conditions is largely the same. Thus, establishment of a definitive diagnosis is important mainly for planning the subacute and long-term management of the patient's disease.

Management

In a life-threatening situation (eg, cardiovascular collapse) or in the setting of frank generalized tetany, 10 mL of calcium gluconate (93 mg of elemental calcium) can be administered intravenously over a 5- to 10-minute period and repeated if necessary. Less acute but recurrent hypocalcemic episodes can be managed with continuous calcium infusions. Nine hundred and thirty milligrams (10 ampules) of calcium gluconate can be mixed in 500 mL of D_5W. Infusion rates are established empirically. The initial infusion rate is typically 0.3 mg/kg/h but may be increased to as much as 2 mg/kg/h in patients with high demands for calcium (eg, hungry bones syndrome). Intravenous calcium should be given through a central vein since extravasation of the infusate can lead to tissue sloughing. Caution should also be exercised in infusing calcium in patients with hyperphosphatemia (precipitation of calcium phosphate salts in soft tissues), hypokalemia, or those on digoxin therapy (increased risk of cardiac arrhythmias).

As the acute situation resolves, selected patients become candidates for chronic replacement therapy. Those with minor hypocalcemia may be treated with calcium supplements alone. Administration of 1 to 3 g of elemental calcium (calcium carbonate contains about 40% elemental calcium by weight) may suffice to restore calcium to the low-normal range and eliminate hypocalcemic symptoms. In those with more refractory hypocalcemia—a

category that includes most patients with hypoparathyroidism—the addition of some form of vitamin D or a derivative is required. If ergocalciferol (vitamin D_2) is employed to manage patients with hypoparathyroidism, supraphysiologic doses (50,000 U/d or higher) may be required, reflecting the limited capacity for 1-α hydroxylation of the vitamin D pro-hormone in the hypoparathyroid state. 1,25-Dihydroxycholecalciferol (calcitriol) works faster than vitamin D_3 and circumvents the 1-α-hydroxylase blockade, but it is more expensive and the risk of acute hypercalcemia may be higher. It is given in a dose of 0.25 to 1 μg daily.

In all cases the goal of therapy should be to alleviate hypocalcemic symptoms and restore serum calcium levels to the low-normal range. This can usually be accomplished with some combination of vitamin D and supplemental calcium. The latter allows the practitioner additional flexibility in controlling serum calcium levels without requiring frequent dose modification of the longer-acting vitamin D. Efforts to push calcium levels higher may come at the expense of significant hypercalciuria and increased risk of renal stone formation. Difficulty in maintaining calcium even in the low-normal range without unacceptable hypercalciuria may be managed by addition of a thiazide diuretic (hydrochlorothiazide, 25-100 mg/d) to the regimen. These agents reduce hypercalciuria and, secondarily, raise serum calcium levels. If thiazides are used, careful follow-up is required to guard against the possibility of iatrogenic hypercalcemia. More refractory patients may be managed with recombinant PTH (1-34) or PTH (1-84) (see Chapter 8). Use of these agents is often successful in restoring normocalcemia and/or reducing doses of more conventional medications used in this setting. Efforts should be made to restore calcium to the target range and maintain it at this level indefinitely. Chronic hypocalcemia is associated with development of subcapsular cataracts and basal ganglia calcification which, in some cases, leads to the development of a parkinsonism-like syndrome.

HYPONATREMIA

Clinical Setting

Hyponatremia is the most frequent electrolyte problem observed in hospitalized patients (see Chapter 5). It is often relatively asymptomatic when mild to moderate in severity and subacute to chronic in its time course of development. However, significant hyponatremia (<120 mEq/dL) of rapid onset is frequently symptomatic and can be life-threatening.

Hyponatremia typically occurs in one of three clinical settings, each of which is linked to a specific pathophysiologic paradigm (Table 24–11). Hypovolemic hyponatremia is associated with volume contraction. As intravascular volume is reduced by more than about 9%, there is a nonosmotic stimulation of antidiuretic hormone (ADH; vasopressin) release as the body attempts to retain water to support intravascular volume. Hyponatremia of this type is seen with protracted vomiting, diarrhea, or excessive sweating, particularly when fluid losses are replenished with water or hypotonic fluids alone. Volume contraction and hyponatremia may also be seen with disorders of renal sodium handling (eg, diuretic

TABLE 24–11   Classification of hyponatremia.

Hypovolemic
Extrarenal volume loss
Excessive perspiration
Losses due to widespread skin involvement (eg, burns)
Vomiting
Diarrhea
Renal volume loss
Diuretic use
Salt-wasting nephropathy
Cerebral salt wasting
Osmotic diuresis with selective repletion of water
Hypervolemic
Congestive heart failure
Nephrotic syndrome
Cirrhosis with ascites
Chronic renal failure
Normovolemic
Syndrome of inappropriate antidiuretic hormone secretion
Intracranial disease
Pulmonary disease
Drugs (eg, opioids)
Severe pain or emotional distress
Nausea and vomiting
Postsurgical
Resetting of the osmostat
Adrenal insufficiency (may be hypovolemic in presence of severe mineralocorticoid deficiency)
Hypothyroidism
Water intoxication
Psychogenic polydipsia
Non-psychogenic polydipsia
Excessive administration of parenteral hypotonic fluids
Post-transurethral prostatectomy

use, mineralocorticoid deficiency, or other salt-wasting syndromes). Urinary Na^+ concentration is typically elevated (>20 mEq/L) in these latter disorders, whereas in the former, urinary Na^+ concentration is low, reflecting aggressive resorption of Na^+ in all tubular segments.

Hypervolemic hyponatremia includes those edematous disorders typified by paradoxical retention of Na^+ and water in the face of a total body excess of each. Specific causes of hyponatremia in this group include congestive heart failure, cirrhosis of the liver with ascites, and nephrotic syndrome. Hyponatremia in this setting presumably results from perceived hypoperfusion by baroreceptors in the arterial and venous circulation. Neural impulses transmitting this information to the hypothalamus trigger an increase in ADH release and net water retention.

Normovolemic hyponatremia is probably the most heterogeneous category and the most difficult to define pathophysiologically. It includes the syndrome of inappropriate ADH secretion (SIADH), hypothyroidism, glucocorticoid insufficiency (eg, secondary adrenal insufficiency), psychogenic polydipsia, postoperative hyponatremia, and hyponatremia seen following transurethral resection of the prostate (see Table 24–11).

Diagnosis

Acute hyponatremia, developing over the course of 24 hours or less, presents with headache, nausea, vomiting, lethargy,

restlessness, hyporeflexia, seizures, and altered sensorium, which may progress to stupor and coma. Hyponatremic encephalopathy is responsible for 30% of new-onset seizures in the ICU. These findings are thought to result from cerebral edema as the hypotonic extracellular compartment shifts water into the cerebral cortical cells. Such fluid shifts are opposed early through a reduction in intracellular electrolyte concentration and later by depletion of intracellular solutes (eg, amino acids). This acts to reduce the osmotic gradient and limit the net movement of fluid into brain. With chronicity, such solute shifts can reduce brain water content to near-normal. Thus, the acuteness of the reduction in serum Na^+ concentration and the magnitude of the reduction are important markers of potential morbidity in this disorder. Young menstruating women are particularly susceptible to the deleterious effects of cerebral edema in the postoperative setting. They are 25 times more likely than postmenopausal women or men to die or have permanent brain damage. This increased susceptibility may reflect the effects of estrogen and progesterone to promote solute accumulation in the cells of the central nervous system. Such accumulation would be predicted to increase the osmotic drive that leads to cerebral edema in these patients. Other patient groups at particular risk include children and patients with hypoxia or hepatic disease.

The first step in making the diagnosis is to exclude the presence of pseudohyponatremia. The latter results from high circulating concentrations of triglycerides or osmotically active solutes (eg, glucose or proteins) in circulating plasma. Hypertriglyceridemia artifactually lowers serum sodium by physically excluding it from the sizable nonaqueous phase of the sample being measured. This is usually readily detected in the laboratory (eg, by noting the presence of lactescent serum) and is corrected by centrifuging the sample prior to measuring Na^+ concentration in the aqueous phase. Osmotically active solutes, like glucose, draw water from the intracellular to the extracellular compartment, where it may transiently lower existing electrolyte (eg, Na^+) concentrations (see Diabetic Ketoacidosis, discussed earlier).

Assuming that the presence of hyponatremia is confirmed, an attempt to examine the different diagnostic possibilities listed earlier should be initiated. Evidence of congestive heart failure, cirrhosis, or nephrotic syndrome is usually apparent on physical examination and confirmable with standard laboratory or imaging studies. Similarly, renal dysfunction should be excluded using conventional renal function tests. Thiazide diuretic use is a frequent cause of hyponatremia and should be investigated early in the evaluation. A careful history of water consumption should be obtained and measurements of water intake in a monitored setting made to exclude psychogenic polydipsia or dipsogenic diabetes insipidus. Hypothyroidism can be excluded with measurement of plasma TSH and free thyroxine levels and glucocorticoid deficiency through an ACTH stimulation test (see Chapters 7 and 9).

Nonosmotic, non-volume-driven ADH secretion is found in SIADH. This is typically a diagnosis of exclusion in non-volume-contracted individuals without evidence of edema, renal insufficiency, hypothyroidism, or adrenal insufficiency. Serum Na^+ and osmolality are low in the face of a concentrated urine (at least >100

mOsm/kg). Urine Na^+ may be modestly elevated (>40 mEq/L with normal dietary NaCl intake), reflecting activation of natriuretic pathways responding to the increase in total body fluid volume. If findings are equivocal, an abnormal water load test (inability to excrete at least 90% of a 20 mL/kg water load in 4 hours or failure to dilute urine osmolality to below 100 mOsm/kg) can be used to confirm the diagnosis. Hypouricemia and low blood urea nitrogen levels are suggestive but not diagnostic of SIADH. Serum uric acid less than 4 mg/dL in the presence of hyponatremia has a positive predictive value for SIADH of greater than 70%. SIADH is seen with a variety of disorders affecting the central nervous system (eg, encephalitis, multiple sclerosis, meningitis, psychosis), the pulmonary system (eg, tuberculosis, pneumonia, aspergillosis), or as a paraneoplastic process associated with a number of solid tumors (eg, small cell carcinoma of the lung, carcinoma of the pancreas, bladder, or prostate). It may also be seen with certain types of drugs (eg, cyclophosphamide, vinca alkaloids, opioids, prostaglandin synthesis inhibitors, tricyclic antidepressants, carbamazepine, clofibrate, serotonin reuptake inhibitors and the recreational drug "ecstasy," 3,4 methylene dioxymethamphetamine) and in clinical situations associated with pain, stress, nausea, vomiting, hypoxemia, hypercapnia, hypoglycemia, and the perioperative state.

Some difficulty may be encountered in differentiating SIADH from a second hyponatremic syndrome called cerebral salt wasting. The latter is also associated with central nervous system disease, particularly subarachnoid hemorrhage. It is thought to be due to a centrally mediated renal wasting of sodium with consequent volume contraction, volume-dependent activation of ADH secretion, and hyponatremia. A recent study that used central venous pressure to separate groups with SIADH versus cerebral salt wasting (CVP <5 cm water defined volume contraction of cerebral salt wasting) in patients with subarachnoid hemorrhage and hyponatremia, 63% were thought to be due to SIADH and only 6.5% to cerebral salt wasting. It has been suggested that atrial natriuretic peptide or brain natriuretic peptide may play a central role in mediating the natriuresis associated with this disorder. Comparison of the features of these two disorders (Table 24–12) suggests that clinical or biochemical evidence of volume contraction is the major way to differentiate cerebral salt wasting from the euvolemic hyponatremia of SIADH. This is an important distinction, because therapy in cerebral salt wasting involves intravascular volume repletion, whereas in SIADH fluid restriction may represent first-line therapy. There is also a relative contraindication to volume depletion in patients with SAH due to the risk of vasospasm. In this instance administration of hypertonic saline (3% NaCl) may represent an effective approach to correction of the hyponatremia, regardless of etiology.

Management

When the primary stimulus provoking water retention (eg, diuretic use) or consumption (eg, psychogenic polydipsia) can be identified, specific therapy represents the most rational approach for long-term management.

TABLE 24–12 Comparison of laboratory findings in SIADH with CSW.

	SIADH	CSW
Intravascular volume contraction[a]	No	Yes
Serum sodium	↓	↓
Urine sodium	↑	↑↑
Urine volume	N or ↓	↑
Endocrine		
Plasma ADH	↑	N or ↑
Plasma aldosterone	↓	N or ↑
Plasma renin	↓	N or ↓
Plasma ANP	↑	↑
Urea nitrogen, serum	N or ↓	↑
Serum uric acid	↓	N or ↑

[a]Based on clinical assessment or direct measurement of plasma volume.

Abbreviations: ADH, antidiuretic hormone; ANP, atrial natriuretic peptide; CSW, cerebral salt wasting; N, normal; SIADH, syndrome of inappropriate antidiuretic hormone secretion.

When the cause of hyponatremia is unclear or unaddressable (eg, SIADH), a more generic approach may be adopted. Patients with asymptomatic (eg, mild or chronic) hyponatremia can be managed with water restriction, athough noncompliance and lack of efficacy when urine osmolality is greater than 500 mOsm/kg or urine-to-plasma Na concentration is greater than 1, limit its usefulness. Calculations of daily water intake should incorporate that included in nonliquid foods consumed by the patient. For patients who are symptomatic, non-acute but unable to adhere to water restriction, treatment with an oral vasopressin antagonist, tolvaptan (15-60 mg/d), or demeclocycline (600-1200 mg/d in divided doses), an antibiotic that uncouples ADH from activation of its receptor, may be sufficient to control serum Na$^+$ levels. Use of tolvaptan is limited to 30 days due to potential hepatotoxicity, and it should not be used in patients with liver disease. Water restriction is not required with demeclocycline therapy and may even be deleterious. Such therapy should be carefully monitored to guard against precipitous dehydration and renal insufficiency. Alternatively, patients can be managed with regular administration of a loop diuretic (eg, furosemide), which leads to excretion of urine approximately half the tonicity of plasma (loop diuretics disrupt the osmotic gradient required for urine concentration). Loop diuretics should be used concomitantly with NaCl supplementation (2-3 g/d) to increase urinary solute excretion and in that way amplify urinary water loss.

Symptomatic acute hyponatremia is an indication for hypertonic saline (3% NaCl) administration. Estimates of excess total body water can be made using the following formula:

$$\text{Excess water} = \frac{295 - \text{Plasma osmolality}}{295} \times 0.6 (\text{Body weight})$$

with plasma osmolality in milliosmoles per kilogram, excess water in liters, and body weight in kilograms.

The rise in serum sodium affected by 1 L of 3% saline infusate can be calculated by the following formula from Adrogue and Madias:

$$\Delta Na^+ = \frac{\text{Influsate } Na^+ - \text{Serum } Na^+}{\text{Total body water} + 1}$$

where ΔNa^+ is the change in sodium concentration effected by administration of 1 L of infusate; infusate Na^+ is the concentration of Na^+ in the infusion (eg, 513 mmol/L for 3% NaCl); and total body water is body weight (kg) × 0.6 (children or nonelderly men), × 0.5 (nonelderly women and elderly men), or × 0.45 (elderly women). The +1 in the denominator accounts for the volume of the infusate. Based on the calculated ΔNa^+, one can adjust the infusion rate to provide the desired increase in serum Na^+ over a fixed time interval. An alternative guide to initial therapy estimates that 1 mL/kg of 3% NaCl will increase serum Na^+ by approximately 1 mEq/L. Infusion rate algorithms are intended only as a guide and should be adjusted, based on actual Na^+ measurements, to raise serum Na^+ levels by no more than 0.5 mEq/h (total <8-12 mEq/d) to diminish the risk of central demyelination (see next). If necessary, infusion rates can be adjusted to increase serum sodium by 1 to 2 mEq/h for short periods of time in symptomatic patients; however, the limitation to 8 to 12 mEq/d should be adhered to, if possible. Furosemide can be administered, if necessary, to avoid intravascular fluid overload; however, given the ability of this drug to promote excretion of a hypotonic urine, the clinician should be aware that the plasma osmolality may increase faster than predicted by the formula set forth earlier. Once serum Na^+ reaches 130 mEq/L, hypertonic saline infusion should be terminated and fluid restriction or normal saline (0.9% NaCl) plus furosemide used for the final correction of serum osmolality.

For moderately symptomatic hyponatremia in a patient requiring parenteral medication, the vasopressin receptor antagonist conivaptan may be used (20 mg IV initially followed by infusion of 20 mg/d for 1 to 3 days; infusion rate may be increased to 40 mg/d if serum sodium fails to increase appropriately). Moderate fluid restriction is recommended to facilitate reaching serum sodium goal.

It is important to mention that interventions directed at the cause of the hyponatremia (vs treatment of the hyponatremia itself) also require close monitoring. Rapid correction of hyponatremia through administration of glucocorticoids in adrenal insufficiency or volume repletion in thiazide-induced hyponatremia, for example, may be associated with central myelinolysis. Careful monitoring of serum Na^+ levels in the post-treatment period is mandatory. If Na^+ appears to be rising too rapidly (>1 mEq/h), administration of hypotonic fluids or a small dose of desmopressin acetate (0.25-1 μg parenterally) may be indicated.

Complications

Central pontine myelinolysis was first described in alcoholics and malnourished patients. In the original descriptions it was

characterized by demyelination confined to the pons, resulting in quadriplegia and, not infrequently, death. Subsequent observational studies linked it to the treatment of hyponatremia. The classic presentation is of a patient who is aggressively treated for hyponatremia with resolution of the presenting findings (ie, those of cerebral edema), only to develop symptoms of mutism, dysphasia, spastic quadriparesis, pseudobulbar palsy, delirium, and, in many cases, death. Surviving patients often have severe neurologic sequelae. More recent studies using CT and MRI indicate that myelinolysis is not confined to the pons but present in many extrapontine locations as well. Lesions are typically symmetrically distributed and clustered in areas where there is close juxtaposition of gray and white matter.

There has been considerable controversy about the prevalence of this particular syndrome and its relationship to the treatment of hyponatremia. However, both animal and human studies are strongly suggestive of a link between this syndrome and rapid, aggressive correction of hyponatremia. Given the imperfect state of our understanding of this disorder, it would seem prudent to approach the correction of chronic hyponatremia, where solute distribution and water content in the brain are undoubtedly altered because of the chronicity of the disorder, cautiously, with rates of correction no greater that 0.5 mEq/h, as indicated earlier. The risk of solute redistribution in documented acute hyponatremia (ie, duration <24 hours) is substantially reduced. Clinical signs of cerebral edema in this setting may be approached more aggressively, although repletion rates greater than 1 mEq/h with a maximal correction of 12 mEq over the first 24 hours should be avoided if possible.

DIABETES INSIPIDUS

Clinical Setting

Diabetes insipidus is a disorder that is due to absolute or relative deficiency in the circulating levels or bioactivity of ADH (vasopressin) (see Chapter 5).

ADH release is normally suppressed at plasma osmolalities below 285 mOsm/kg, leading to generation of a maximally dilute urine (<100 mOsm/kg). Above 285 mOsm/kg, ADH secretion increases linearly with plasma osmolality. At a plasma ADH level of 5 pg/mL, which corresponds to a plasma osmolality of about 295 mOsm/kg, the urine is maximally concentrated. Despite the fact that plasma ADH levels continue to rise beyond this point, there is no further concentration of the urine. At approximately 295 mOsm/kg, the osmotic threshold for the thirst mechanism is activated. Thirst also rises linearly with increasing plasma osmolality. It provides the body's main defense against hypertonicity, much as suppression of ADH secretion guards against hypotonicity. Provided that the patient remains awake and able to drink and provided that the thirst mechanism remains intact, even complete deficiency of ADH secretion can be adequately compensated by increased water intake.

Defects in ADH secretion (central diabetes insipidus) may occur on a heritable basis, occasionally in association with diabetes mellitus, optic atrophy, and sensorineural deafness (DIDMOAD,

or Wolfram syndrome). It may also be seen as a component of a polyendocrine autoimmune syndrome (see Chapter 2). Secondary causes of central diabetes insipidus include head injury (often with stalk section), pituitary surgery, granulomatous disease (sarcoidosis, tuberculosis, histiocytosis X), infections, vascular aneurysms and thrombosis, and tumors (craniopharyngioma, dysgerminoma, meningioma, metastatic disease from breast, lung, or gastrointestinal tumors).

Osmoreceptors for ADH release are believed to reside in the organum vasculosum of the lamina terminalis, whereas those controlling thirst are believed to lie in an independent but neighboring location. Thus, those patients with an isolated defect in ADH secretion (either in the osmoreceptors or the secretory nuclei) are protected from severe dehydration by activation of the thirst mechanism. In those with combined involvement of both ADH secretion and the osmoreceptors controlling thirst (hypodipsic or adipsic diabetes insipidus), the risk of volume contraction and severe dehydration is extreme (discussed later).

Adequacy of the renal response to ADH requires adequate delivery of glomerular filtrate to distal tubular segments, adequacy of tubular function in the ascending limb of the loop of Henle (to establish and maintain the gradient of medullary tonicity and to generate free water for excretion), and a normal response to vasopressin (ie, intact signal transduction mechanism) in the collecting duct. Unresponsiveness to ADH can be due to genetic lesions in the V_2 receptor (X-linked recessive diabetes insipidus) or the functionally linked aquaporin-2 water channel (autosomal recessive diabetes insipidus). It may also be seen in an acquired form with hypokalemia or hypercalcemia, in various forms of intrinsic renal disease (eg, medullary cystic disease) and following therapy with demeclocycline or lithium.

Diagnosis

The presentation of diabetes insipidus in an alert, conscious patient is typically an abrupt onset of polyuria and polydipsia. Hypertonicity is avoided as long as the thirst mechanism remains intact and water intake is able to keep up with urinary losses. In a patient who is unconscious or otherwise incapable of communicating the need for fluids or in patients with coexistent adipsia, polyuria is transient and quickly followed by evidence of severe dehydration and hyperosmolality. Urine volumes in this setting may be normal or even reduced. The clinical findings in hypernatremia associated with diabetes insipidus are dominated by the effects of cellular dehydration and contraction of intravascular volume. In the brain, this can lead to increased traction on dural veins and venous sinuses. This may result in avulsion of vessels from their cranial attachments and intracranial hemorrhage. Other findings include irritability, lethargy, weakness, muscle twitching, hyperreflexia, seizures, and coma.

The diagnosis of polyuria is usually reserved for urine outputs greater than 2.5 L/d. A careful history should be taken to document oral fluid intake (particularly beer or other hypotonic fluids) or parenteral fluid administration (eg, in a postoperative setting). Neurologic or endocrine symptoms that might suggest either a hypothalamic or intrasellar mass, use of medications that could

impair water reabsorption (eg, furosemide, demeclocycline, or lithium) or intercurrent conditions that might mimic diabetes insipidus (eg, osmotic diuresis associated with diabetes mellitus or resolving obstructive uropathy) should also be investigated.

Initial evaluation should include measurement of serum sodium, plasma osmolality, and urine osmolality. Diabetes insipidus is typically associated with a high-normal or elevated serum sodium and osmolality in the face of submaximally concentrated urine. Plasma glucose measurement and standard renal function tests should help exclude osmotic diuresis as contributing to the polyuria. Measurement of serum K^+ and Ca^{2+} assists in excluding polyuria resulting from hypokalemia and hypercalcemia, respectively.

In patients with equivocal initial tests, a water deprivation test may be performed. This involves serial measurements of serum Na^+, urine volume, and osmolality in the absence of fluid intake until serum Na^+ is greater than 146 mmol/L or urine osmolality plateaus (<10% change with three successive determinations), and the patient has lost more than 2% of body weight. The test should be stopped if the decrease in body weight reaches 3% or the patient demonstrates orthostatic hypotension. Patients with diabetes insipidus become progressively more hypertonic with water deprivation but fail to increase urine osmolality. Administration of desmopressin promotes water retention and an increase in urine osmolality in patients with central but not nephrogenic diabetes insipidus. Increases in osmolality of more than 50% indicate central diabetes insipidus, whereas responses of less than 10% suggest nephrogenic diabetes insipidus. Responses between 10% and 50% are equivocal. Plasma ADH levels may be of assistance in defining the nature of the diabetes insipidus in the latter setting. In the presence of high plasma but low urine osmolality, increased plasma ADH levels are associated with nephrogenic diabetes insipidus, whereas subnormal levels are found with central diabetes insipidus. It may be difficult to establish the diagnosis of mild partial central diabetes insipidus. These patients can often achieve maximally concentrated urine during extreme dehydration, reflecting reduced GFR, and stimulated vasopressin secretion; however, as plasma osmolality is corrected and intravascular volume is restored, they typically begin to dilute their urine at a plasma osmolality that is well above normal.

Coexistence of glucocorticoid deficiency may mask the presence of central diabetes insipidus. The development of polyuria with the initiation of glucocorticoid replacement should alert the clinician to the possible presence of diabetes insipidus.

In the 48 hours following pituitary surgery or head trauma, transient diabetes insipidus with polyuria is not uncommon. Over the ensuing 2 to 14 days, a period of antidiuresis and hyponatremia may dominate the clinical course. This in turn may be followed by persistent polyuria. The first phase is thought to result from transient dysfunction or stunning of the ADH-producing neurons of the hypothalamus, the second from leakage of ADH from damaged or dying neurons, and the third from permanent loss of ADH-secretory neurons. Not all patients progress through the entire series of events. It is important to recognize the existence of this syndrome and the natural progression of the disorder in evaluating the need for chronic therapeutic intervention.

In pregnancy, there is a resetting of the osmostats controlling both ADH and thirst, resulting in a state of physiologic hypo-osmolality (about 10 mOsm/kg below that seen in the nonpregnant state). In addition, elevations in placental vasopressinase may rarely promote polyuria in patients with otherwise compensated partial central diabetes insipidus. The appropriate treatment for this disorder is desmopressin acetate, an ADH analog that is resistant to degradation by vasopressinase.

Virtually anything that increases the rate of tubular flow (eg, primary polydipsia or central diabetes insipidus) may create a state of functional nephrogenic diabetes insipidus. This is due to washout of the medullary tonicity responsible for generating the osmotic gradient that promotes movement of water from the tubular lumen into the medullary interstitium as well as down regulation of aquaporin-2 water channels as a result of vasopressin deficiency. A therapeutic trial of 1-deamino-8-D arginine vasopressin (DDAVP) for 1 to 2 days often restores medullary tonicity and facilitates the diagnostic workup. Primary polydipsia (eg, psychogenic polydipsia or dipsogenic diabetes insipidus) presents with a dilute urine, but plasma osmolality and serum Na^+ levels are typically low or low-normal and ADH levels are low. Moreover, due to the elevation in tubular flow rates, the response to desmopressin may be limited in magnitude (<10%, which is less than that predicted for central diabetes insipidus). This may make it difficult to differentiate primary polydipsia from partial forms of nephrogenic diabetes insipidus. Plasma ADH levels may be of assistance in making this distinction (elevated in nephrogenic diabetes insipidus and suppressed in psychogenic polydipsia).

Management

Acute diabetes insipidus is characterized by polyuria in the presence or absence of plasma hyperosmolality. As noted earlier, the presence of hyperosmolality is determined by the adequacy of the patient's thirst mechanism. When a hyperosmolar state with clinical evidence of severe dehydration is present, it is necessary to correct intravascular volume. This is accomplished by administering hypotonic fluids (eg, D_5W), which restore intravascular volume and move plasma osmolality back toward normal. The rate at which the hyperosmolality is corrected is dictated in part by the severity of the clinical symptoms and by the chronicity of the disorder. Chronic hyperosmolality (>24 hours) leads to the accumulation of idiogenic osmoles (eg, taurine and myoinositol) in the central nervous system that serve to offset the osmotically driven movement of water out of that compartment. Aggressive correction of the hyperosmolar state can lead to cerebral edema if the rate of hypotonic fluid administration greatly exceeds the rate at which the neurons in the central nervous system eliminate these idiogenic osmoles. If severe volume contraction is present and blood pressure is reduced, intravascular volume should be repleted initially with normal saline (0.9% NaCl). Following expansion of intravascular volume with normal saline, correction of plasma osmolality

with D5W at a rate of 0.5 to 1 mOsm/h—not to exceed 15 mOsm within the first 24-hour period—appears to be reasonably effective in reducing plasma tonicity without incurring an increased risk of cerebral edema. Typically, higher infusion rates are used early in the course when serum osmolalities are highest and reduced as osmolality decreases into the range of 330 mOsm/kg. Correction of the total free water deficit should be spread over about 48 hours. Total water deficit is defined earlier in the section dealing with DKA. Calculation of hypotonic fluid infusion rates can be made using the formula for change in serum sodium set out in the preceding section, substituting 0 mmol/L for D_5W, 77 mmol/L for 0.45% NaCl, or 154 mmol/L for 0.9% NaCl into the equation, as needed, to calculate the net reduction in serum Na^+. Again, based on the calculated ΔNa^+, one can adjust the rate of delivery of the infusate to provide a specific increment in serum Na^+ concentration over a defined time interval. If signs of cerebral edema (see Hyponatremia, discussed earlier) appear, hypotonic fluids should be discontinued and appropriate administration of hypertonic fluids (eg, mannitol) initiated.

If the diagnosis of central diabetes insipidus has been established, the patient should receive parenteral desmopressin acetate. The initial dose, in the range of 1 to 2 μg every 24 hours (administered intravenously or intramuscularly), should be given in the evening with the aim of controlling nocturia and maintaining daily urine output under 2 L. If polyuria returns well before the end of the 24-hour dosing interval, the single dose can be increased or split doses can be administered every 12 hours. Ideally, one would like to see some degree of breakthrough polyuria at the end of the dosing interval. This serves to guard against the development of iatrogenic hyponatremia. If this proves difficult to achieve, a dose can be skipped every 48 to 72 hours. Once the patient has been stabilized on a fixed dose of parenteral desmopressin, conversion to a formulation more suitable for the outpatient setting is indicated. Desmopressin acetate is available in liquid form for insufflation through a nasal cannula or as a fixed-dose (10 μg) nasal spray. The former has the advantage of greater flexibility in adjusting the nasal dose, whereas the latter offers greater convenience for some patients. Parenteral desmopressin is generally about 10 times more potent than the nasally administered drug; however, intranasal doses should be titrated for each individual patient. Desmopressin is also available in an oral form. Effective doses can range widely (50-1200 μg/d in divided doses) and should be adjusted for the individual patient.

Treatment of nephrogenic diabetes insipidus is more complex because the problem is one of resistance to—rather than deficiency of—the endogenous hormone. If possible, one should attempt to address the underlying cause of ADH resistance. This can be accomplished by correcting electrolyte abnormalities (eg, hypokalemia or hypercalcemia) or discontinuing medications (eg, demeclocycline, lithium) that are likely to contribute to ADH insensitivity. Failing this, administration of a thiazide diuretic together with salt restriction often reduces the polyuria. This is assumed to result from contraction of intravascular volume with increased proximal reabsorption of fluids and solutes and, consequently, limited availability of fluid in distal nephron

segments for free water generation and excretion; however, there may be a second, ADH-independent effect on the aquaporin-2 channels in the collecting duct. Prostaglandins are endogenous antagonists of ADH activity in the collecting duct. Administration of cyclooxygenase inhibitors (eg, indomethacin, 100 mg/d in divided doses) may improve sensitivity to endogenous ADH and reduce polyuria. If the impairment in ADH responsiveness is mild, higher doses of synthetic desmopressin (DDAVP) occasionally prove effective in promoting water retention. Amiloride is the preferred treatment for lithium toxicity. It is thought to prevent the uptake of lithium in collecting duct cells. However, lithium-induced diabetes insipidus may not abate following discontinuation of the drug.

Disorders of thirst in the setting of central diabetes insipidus deserve special mention with regard to therapy. Excessive thirst due to altered osmoreceptor function or behavioral conditioning (eg, before desmopressin treatment) may result in severe hyponatremia once desmopressin treatment is initiated. The patient should be cautioned about excessive consumption of water, and fluid must be restricted if necessary.

Adipsic diabetes insipidus is probably one of the most difficult therapeutic problems faced by endocrinologists. Affected patients have essentially lost all ability to regulate water metabolism on their own. This function must be taken over by the medical team providing their care. Desmopressin is administered in a fixed parenteral dose that is sufficient to reduce urine output to 1.5 to 2 L/d with a fluid intake of 2 to 2.5 L. The difference here is intended to cover the daily insensible losses (about 500-1000 mL) and may need to be adjusted empirically to maintain normal fluid homeostasis. After plasma osmolality and sodium are in the normal range, fluid intake is balanced against urine output. Any net change in urine output over an 8-hour period (this may be extended as the patient's fluid requirements become more predictable) is accommodated by modifying the fluid orders for the ensuing 8 hours. The patient should also be weighed daily. Any alteration in weight is usually reflective of net changes in water retention and can be corrected through modification of fluid intake. Plasma osmolality and serum sodium should be monitored once or twice weekly and appropriate changes in fluid administration made depending on the direction and magnitude of the shift in osmolality. With this system of redundant monitoring, urine volumes and plasma osmolality can be reasonably well controlled for protracted periods of time.

Complications

If left untreated, polyuria secondary to diabetes insipidus can lead to dilation of the collecting system, hydronephrosis, and renal dysfunction. This is the dominant reason for treating polyuria even in patients with an intact thirst mechanism who are capable of controlling plasma osmolality through water ingestion.

Overly rapid correction of the hyperosmolality can result in cerebral edema. The relative risk of this complication should be minimized by careful attention to the rate at which the water deficit is corrected (discussed earlier).

REFERENCES

Myxedema Coma

Dutta P, Bhansali A, Masoodi SR, et al. Predictors of outcome in myxoedema coma: a study from a tertiary care center. *Crit Care*. 2008;12:R1. [PMID: 18173846]

Kwaku MP, Burman KD. Myxedema coma. *J Intensiv Care Med*. 2007;22:224. [PMID: 1121935]

Sarlis NJ, Gourgiotis L. Thyroid emergencies. *Rev Endocr Metab Disord*. 2003;4:129. [PMID: 12766540]

Yamamoto T, Fukuyama J, Fujiyoshi A. Factors associated with mortality of myxedema coma: report of eight cases and literature survey. *Thyroid*. 1999;9:1167. [PMID: 10646654]

Thyroid Storm

Angell TE, Lechner MG, Nguyen CT, et al. Clinical features and hospital outcomes in thyroid storm: a retrospective cohort study. *J Clin Endocrinol Metab*. 2015;100:451. [PMID: 25343237]

Klubo-Gwiezdzinska J, Wartofsky L. Thyroid emergencies. *Med Clin N Am*. 2012;96:385. [PMID: 22443982]

Scholz GH, Hagemann E, Arkenau C, et al. Is there a place for thyroidectomy in older patients with thyrotoxic storm and cardiorespiratory failure. *Thyroid*. 2003;13:933. [PMID: 14611702]

Yeoung SC, Go R, Balasubramanyam A. Rectal administration of iodide and propylthiouracil in the treatment of thyroid storm. *Thyroid*. 1995;5:403. [PMID: 8563481]

Thyrotoxic Periodic Paralysis

Falhammar H, Thoren M, Galissendorff S. Thyrotoxic periodic paralysis: clinical and molecular aspects. *Endocrine*. 2013;43:274. [PMID: 22918841]

Kung AWC. Thyrotoxic periodic paralysis: a diagnostic challenge. *J Clin Endocrinol Metab*. 2006;91:2490. [PMID: 16608889]

Lin SH, Lin YF. Propanolol rapidly reverses paralysis, hypokalemia and hypophosphatemia in thyrotoxic periodic paralysis. *Am J Kidney Dis*. 2001;37:620. [PMID: 11228188]

Lu K-C, Hsu YJ, Chiu JS, et al. Effects of potassium supplementation on the recovery of thyrotoxic periodic paralysis. *Am J Emerg Med*. 2004;22:544. [PMID: 15666258]

Ryan DP, da Silva MR, Soong TW, et al. Mutations in potassium channel Kir2.6 cause susceptibility to thyrotoxic hypokalemic periodic paralysis. *Cell*. 2010;140:88. [PMID: 20074522]

Tassone H, Mouslin A, Henderson SO. The pitfalls of potassium replacement in thyrotoxic periodic paralysis: a case report and review of the literature. *J Emerg Med*. 2004;26:157. [PMID: 1490336]

Amiodarone-Induced Hyperthyroidism

Bartelena L, Wiersinga WM, Tanda ML, et al. Diagnosis and management of amiodarone-induced thyrotoxicosis in Europe: results of an international survey among members of the European Thyroid Association. *Clin Endocrinol*. 2004;61:494. [PMID: 15473883]

Basaria S, Cooper DS. Amiodarone and the thyroid. *Am J Med*. 2005;118:706. [PMID: 15989900]

Conen D, Melly L, Kaufmann C, et al. Amiodarone-induced thyrotoxicosis. *J Am Coll Cardiol*. 2007;49:2350. [PMID: 17572251]

Eaton SE, Euinton HA, Newman CM, et al. Clinical experience of amiodarone-induced thyrotoxicosis over a 3-year period: role of colour-flow Doppler sonography. *Clin Endocrinol (Oxf)*. 2002;56:33. [PMID: 11849244]

Acute Adrenal Insufficiency

Arafah BM. Hypothalamic-pituitary-adrenal function during critical illness: limitations of current assessment methods. *J Clin Endocrinol Metab*. 2006;91:3725. [PMID: 16882746]

Cooper MS, Stewart PM. Corticosteroid insufficiency in acutely ill patients. *N Engl J Med*. 2003;348:727. [PMID 12594318]

Dorin RI, Qualls CR, Crapo LM. Diagnosis of adrenal insufficiency. *Ann Intern Med*. 2003;139:194. [PMID: 12899587]

Hamarahian AH, Oseni TS, Arafah B. Measurements of serum free cortisol in critically ill patients. *N Engl J Med*. 2004;350:1629. [PMID: 15084695]

White K, Arlt W. Adrenal crisis in treated Addison's disease: a predictable but under-managed event. *Eur J Endocrinol*. 2010;162:115. [PMID: 19776201]

Pituitary Apoplexy

Briet C, Salenaves S, Chanson P. Pituitary apoplexy. *Endocrinol Metab Clin N Am*. 2015;44:199. [PMID: 25732655]

Johnston PC, Hamrahian AH, Weil RS, et al. Pituitary tumor apoplexy. *J Clin Neuroscience*. 2015;22:939. [PMID 25800143]

Nawar RN, AbdelMannan D, Selman WR, Arafah BM. Pituitary tumor apoplexy: a review. *J Intensive Care Med*. 2008;23:75. [PMID: 18372348]

Diabetic Ketoacidosis

Fleckman AM. Diabetic ketoacidosis. *Endocrinol Metab Clin North Am*. 1993;22:181. [PMID: 8325282]

Hom J, Sinert R. Is fluid therapy associated with cerebral edema in children with diabetic ketoacidosis? *Ann Emer Med*. 2008;52:69. [PMID: 18387706]

Kamel KS, Halperin ML. Acid-base problems in diabetic ketoacidosis. *N Engl J Med*. 2015;372:546. [PMID: 25651248]

Kitabchi AE, Umpierrez GE, Fisher JN, et al. Thirty years of personal experience in hyperglycemic crises: diabetic ketoacidosis and hyperglycemic hyperosmolar coma. *J Clin Endocrinol Metab*. 2008;93:1541. [PMID: 18270259]

Kitabchi AE, Umpierrez GE, Miles JM, Fisher JN. Hyperglycemic crises in adult patients with diabetes. *Diabetes Care*. 2009;32:1335. [PMID: 19564476]

Nyenwe EA, Kitabchi AE. Evidence-based management of hyperglycemic emergencies in diabetes mellitus. *Diab Res Clin Practice*. 2011;94:340. [PMID: 21978840]

Nyenwe EA, Razavi LN, Kitabchi AE, et al. Acidosis: the prime determinant of depressed sensorium in diabetic ketoacidosis. *Diab Care*. 2010;33:1837. [PMID: 20484127]

Wolfsdorf J, Glaser N, Sperling MA. Diabetic ketoacidosis in infants, children and adolescents. *Diabetes Care*. 2006;29:1150. [PMID: 16644656]

Hyperosmolar Nonketotic Coma

Maletkovic J, Drexler A. Diabetic ketoacidosis and hyperglycemic hyperosmolar state. *Endocrinol Metab Clin N Am*. 2013;42:677. [PMID: 24286946]

Pasquel FJ, Umpierrez GE. Hyperosmolar hyperglycemic state: a historic review of the clinical presentation, diagnosis and treatment. *Diabetes Care*. 2014;37:3124. [PMID: 25342831]

Hypocalcemia

Cooper MS, Gittoes NJL. Diagnosis and management of hypocalcemia. *Br Med J*. 2008;336:1298. [PMID: 18535072]

Dickerson RN. Treatment of hypocalcemia in critical illness. *Nutrition*. 2007;23:358,436. [PMID: 17400132, 17360159]

Hypercalcemic Crisis

Berenson JR. Treatment of hypercalcemia of malignancy with bisphosphonates. *Sem Oncol*. 2002;30(2 suppl 5):12. [PMID: 12584690]

Mirrakhimov AE. Hypercalcemia of malignancy: an update on pathogenesis and management. *N Am J Med Sci*. 2015;7(11):483. [PMID: 26713296]

Hyponatremia

Adrogue HJ, Madias NE. Hyponatremia. *N Engl J Med*. 2000;342:1581. [PMID: 10824078]

Berl T. Vasopressin antagonists. *N Engl J Med*. 2015;372:2207. [PMID: 26039601]

Ellison DH, Berl T. The syndrome of inappropriate antidiuresis. *N Engl J Med*. 2007;356:2064. [PMID: 17507705]

Ghali J, Koren MJ, Taylor JR, et al. Efficacy and safety of oral conivaptan: a V1A/V2 vasopressin receptor antagonist, assessed in a randomized, placebo-controlled trial in patients with euvolemic or hypervolemic hyponatremia. *J Clin Endocrinol Metab*. 2006;91:2145. [PMID: 16522696]

Hannon MJ, Behan LA, O'Brien MMC, et al. Hyponatremia following mild/moderate subarachnoid hemorrhage is due to SIAD and glucocorticoid deficiency and not cerebral salt wasting. *J Clin Endocrinol Metab*. 2014;99:291. [PMID: 24248182]

Moritz ML, Ayus JC. Maintenance intravenous fluids in acutely ill patients. *N Engl J Med*. 2015;373:1350. [PMID: 26422725]

Sahun M, Villabona C, Rosel P, et al. Water metabolism disturbances at different stages of primary thyroid failure. *J Endocrinol*. 2001;168:435. [PMID: 11241175]

Singh S, Bohn D, Carlotti AP, et al. Cerebral salt wasting: truths, fallacies, theories and challenges. *Crit Care Med*. 2002;30:2575. [PMID: 12441772]

Sterns R. Disorders of plasma sodium—causes, consequences and correction. *N Engl J Med*. 2015;372:55. [PMID: 25551526]

Diabetes Insipidus

Adrogue HJ, Madias NE. Hypernatremia. *N Engl J Med*. 2000;342:1493. [PMID: 10816188]

Knepper MA, Kwon T-H, Nielsen S. Molecular physiology of water balance. *N Engl J Med*. 2015;372:1349. [PMID: 25830425]

Mavrakis AN, Tritos NA. Diabetes insipidus with deficient thirst: report of a patient and review of the literature. *Am J Kid Dis*. 2008;51:851. [PMID: 18436097]

Morello J-P, Bichet DG. Nephrogenic diabetes insipidus. *Annu Rev Physiol*. 2001;63:607. [PMID: 11181969]

Oiso Y, Rogertson GL, Norgaard JP, et al. Treatment of neurohypophyseal diabetes insipidus. *J Clin Endocrinol Metab*. 2013;98:3958. [PMID: 23884783]

This page intentionally left blank

AIDS Endocrinopathies

Carl Grunfeld, MD, PhD

ACTH	Adrenocorticotropic hormone		**MRI**	Magnetic resonance imaging
AIDS	Acquired immunodeficiency syndrome		**NRTI**	Nucleoside reverse transcriptase inhibitor
BMD	Bone mineral density		**NNRTI**	Non-nucleoside reverse transcriptase inhibitor
CMV	Cytomegalovirus		**PI**	Protease inhibitor
CRH	Corticotropin-releasing hormone		**PPARγ**	Peroxisome proliferator-activated receptor γ
CVD	Cardiovascular disease		**PTH**	Parathyroid hormone
DHEA	Dehydroepiandrosterone		**RANK**	Receptor activator of nuclear factor kappa
DXA	Dual energy x-ray absorptiometry		**REE**	Resting energy expenditure
FSH	Follicle-stimulating hormone		**RXR**	Retinoid X receptor
GH	Growth hormone		**SAT**	Subcutaneous adipose tissue
GHRH	Growth hormone-releasing hormone		**SREBP1**	Sterol regulatory enhancer-binding protein 1
GnRH	Gonadotropin-releasing hormone		T_3	Triiodothyronine
HAART	Highly active antiretroviral therapy		T_4	Thyroxine
HDL	High-density lipoprotein		**TBG**	Thyroxine-binding globulin
HIV	Human immunodeficiency virus		**TNF**	Tumor necrosis factor
IGF	Insulin-like growth factor		**TRH**	Thyrotropin-releasing hormone
IL	Interleukin		**VAT**	Visceral adipose tissue
IMT	Intima media thickness		**VLDL**	Very low density lipoprotein
LDL	Low-density lipoprotein			
LH	Luteinizing hormone			
M/I	Glucose disposal adjusted for insulin level			

Symptoms consistent with endocrine disorders and alterations in endocrine laboratory values are not unusual in individuals infected with the human immunodeficiency virus (HIV). Some of these changes are common to any significant systemic illness; others appear to be more specific to HIV infection or its therapies. Alterations can be found even before clinically significant immunocompromise occurs. As the infected individual becomes immunocompromised, particularly with development of the acquired immunodeficiency syndrome (AIDS), opportunistic infections and neoplasms—as well as the agents used in the treatment of these disorders—can give rise to further changes in endocrine function. Finally, hepatitis C (HCV) coinfection occurs in approximately 1/3 of HIV-infected patients; HCV also induces

changes. Herein we discuss alterations in endocrine function that can accompany HIV/AIDS, focusing on evaluation and interpretation of clinical and laboratory findings.

THYROID DISORDERS

In the era of highly active antiretroviral therapy (HAART), clinical thyroid dysfunction is relatively uncommon in stable HIV-infected patients. In several large studies, the prevalence of hypothyroidism was 1% to 2.5% and hyperthyroidism was 0.5% to 1%. The prevalence of subclinical disease was higher, with subclinical hypothyroidism between 3.5% and 20% and subclinical

hyperthyroidism less than 1%; definitions varied between studies. These data do not support screening for thyroid disease above standard guidelines. With advanced HIV disease, alterations in thyroid function tests do occur, but generally do not result in clinical dysfunction. In patients with AIDS, the effects of opportunistic infections and neoplastic involvement of the thyroid, as well as the effects of some medications used to treat HIV-infected patients, should be considered.

Alterations in Thyroid Function Tests

HIV-infected patients can show alterations in thyroid function tests that are largely asymptomatic. Some of the changes are similar to those seen in the classic euthyroid sick syndrome, whereas others are unique to HIV. Advanced HIV is associated with decrease in thyroid hormone levels, triiodothyronine (T_3) and thyroxine (T_4), similar to that seen in the euthyroid-sick syndrome. When HIV-infected patients were stratified by weight loss and the presence of secondary infection, a decline in T_3 levels corresponded to the severity of disease, consistent with the euthyroid sick syndrome. Deterioration in nutritional status can also contribute to the decrease in T_3 and T_4, because there is a strong correlation between albumin levels and both free T_4 and total T_3 levels. In the euthyroid sick syndrome, there is impaired peripheral T_4 to T_3 conversion by 5' deiodinase. In HIV-infected patients, decreases in free T_4 and free T_3 may be due to decreased extrathyroidal conversion of T_4 to T_3, an increase in serum binding proteins, and/or decreased secretion of TSH.

Despite the similarities in serum thyroid hormone levels between HIV and the euthyroid sick syndrome, significantly ill HIV-infected patients often do not demonstrate the elevated reverse T_3 levels characteristic of the euthyroid sick syndrome. The significance of this difference is unknown.

Thyroxine-binding globulin (TBG) levels are increased in HIV-infected patients, rising progressively with advancing immunosuppression. TBG levels correlate inversely with the CD4 lymphocyte count. The increase in TBG does not appear to be due to generalized changes in protein synthesis, increases in sialylation and altered clearance of TBG, or changes in estrogen levels. The significance of the increased TBG found in HIV-infected patients is unknown. However, increases in TBG affect total T_4 and T_3 measurements, which should be considered when interpreting these tests in HIV-infected patients.

Subtle alterations in TSH dynamics have been reported in stable HIV-infected patients. Despite normal TSH and free T_4 levels, these individuals demonstrate significantly higher TSH values and lower free T_4 values than uninfected controls. In circadian studies, HIV-infected individuals have higher TSH pulse amplitudes with unchanged pulse frequency as well as a higher peak TSH in response to thyrotropin-releasing hormone (TRH) stimulation. These studies are consistent with a subtle state of compensated hypothyroidism; the mechanisms underlying these alterations have not been elucidated.

The typical pattern of alterations in thyroid function tests in HIV-infected patients is outlined in Table 25–1.

TABLE 25–1 Thyroid function tests in HIV-infected patients.

	Basal	After TRH Stimulation
T_3	↓	
T_4	Normal	
Reverse T_3	Normal or ↓	
TBG	↑	
TSH	Normal	↑ Pulse amplitude

↓, decreased; ↑, increased.

Opportunistic Infections and Neoplasms

Opportunistic pathogens and neoplasms can invade the thyroid in HIV-infected individuals but generally do not cause clinical thyroid dysfunction. Autopsy of 100 AIDS patients prior to the era of HAART showed the presence of *Mycobacterium tuberculosis* in 23% of patients' thyroids, followed by cytomegalovirus (CMV) (17%), *Cryptococcus neoformans* (5%), *Mycobacterium avium* (5%), *Pneumocystis carinii* (4%), and other bacteria or fungi (7%).

A few of these pathogens are noteworthy because they have been associated on occasion with both hyper- and hypothyroidism. *P. carinii* has been associated with inflammatory thyroiditis accompanied by hypothyroidism in seven cases, hyperthyroidism in three cases and normal thyroid function in one case. Antithyroid antibodies were negative in all six cases in which they were measured. Radionuclide scanning in seven cases revealed poor visualization of the entire thyroid gland in patients with bilateral disease and nonvisualization of the affected lobe in patients with unilateral disease. Two patients with hyperthyroidism had normalization of thyroid function after treatment of the *P. carinii* infection. Kaposi's sarcoma has also been reported to infiltrate the thyroid gland and resulted in significant destruction and hypothyroidism in at least one case. In two cases, lymphoma was associated with thyroid infiltration, causing thyroidal enlargement. HCV coinfection may increase the prevalence of thyroid disease and perhaps cancer.

Medication Effects

Several medications used to treat HIV-infected patients (eg, rifampin, phenytoin, ketoconazole, and ritonavir) can alter the clearance of thyroid hormone by inducing hepatic microsomal enzymes. Patients with normal thyroid function should not be clinically affected, although decreases in T_4 may be observed. However, patients receiving L-thyroxine may require increased doses, and patients with decreased pituitary or thyroid reserve may develop clinically apparent hypothyroidism. Isolated, well documented reports of need for increased thyroxine doses after starting ritonavir and decreased doses after indinavir have been published; these medications affect glucuronidation, but given the common use of high doses of these drugs in the past it is surprising how rarely problems present. Interferon-alpha, a therapy used for

TABLE 25–2 Medications used in HIV-infected patients that can affect the endocrine system.

Thyroid	**Pancreas-glucose**
Rifampin	Indinavir
Phenytoin	Full dose ritonavir
Ketoconazole	Possibly lopinavir/ritonavir
Possibly ritonavir	Pentamidine
Possibly indinavir	Trimethoprim-
Interferon-alpha	sulfamethoxazole
Adrenals, electrolytes	Dedeoxyinosine (ddI)
Rifampin	Dedeoxycytosine (ddC)
Ketoconazole	Megestrol acetate
Ritonavir	Growth hormone
Megestrol acetate	**Lipids**
Trimethoprim	Ritonavir
Pentamidine	Efavirenz
Sulfonamides	Nevirapine
Amphotericin B	Stavudine
Foscarnet	Tenofovir
Gonads	
Ketoconazole	
Megestrol acetate	
Opiates and Methadone	
Bone, calcium	
Tenofovir	
Adefovir	
Cidofovir	
Foscarnet	
Pentamidine	
Trimethoprim-sulfamethoxazole	
Ketoconazole	
Rifampin, rifabutin	

Kaposi's sarcoma, has been associated with autoimmune diseases of the thyroid, including hyperthyroidism and hypothyroidism. Medications used in HIV-infected patients that can affect the endocrine system are listed in Table 25–2. Nelfinavir inhibits RET and has *in vitro* activity against medullary thyroid cancer. With the advent of HAART, there have been case report series of newly diagnosed autoimmune diseases such as Graves disease, Hashimoto's thyroiditis, and alopecia areata. Immune reconstitution with HAART raises the possibility of subsequent induction of autoimmune diseases. The cases appeared late after immune reconstitution (8-32 months). Possible theories include thymic regeneration or peripheral T lymphocyte expansion causing irregularities in tolerance, leading to autoimmune dysfunction. Other studies, however, did not link autoimmune thyroid disease to HAART, and the low prevalence in large studies suggests that increased screening above standard guidelines may not be warranted.

ADRENAL DISORDERS

Opportunistic infections commonly involve the adrenal glands but rarely occupy enough of the gland to cause adrenal insufficiency. Impaired adrenal reserve without overt symptoms of adrenal insufficiency has been described in patients with HIV. There is controversy over when to give baseline replacement glucocorticoids (or mineralocorticoids) or increased doses in

stressful situations to patients with a subnormal response to dynamic testing, especially when baseline levels are normal or elevated. With restoration to health by treatment of opportunistic infections or HIV, most of these patients no longer have adrenal insufficiency.

Opportunistic Infections and Neoplasms

Opportunistic organisms are found commonly in the adrenal glands of patients dying of AIDS although they rarely cause clinical adrenal insufficiency. CMV has been associated with necrosis of the adrenal gland, although the amount of tissue affected rarely reaches the 90% thought necessary to cause clinical adrenal insufficiency. In one study, the presence of CMV retinitis was associated with an increased rate of adrenal insufficiency compared with AIDS patients without retinitis. Less common opportunistic infections involving the adrenals include *Mycobacterium tuberculosis, Mycobacterium avium-intracellulare, Cryptococcus neoformans, Histoplasma capsulatum, P. carinii,* and *Toxoplasma gondii.* In addition, Kaposi's sarcoma and lymphoma can involve the adrenals, but rarely induce adrenal insufficiency.

Glucocorticoids

Classic clinical symptoms of adrenal insufficiency are seldom seen, but some clinicians view the weakness and weight loss observed in patients with AIDS as an indicator of adrenal insufficiency. HIV-infected patients usually have normal or, even more commonly, elevated basal cortisol levels, which occur most commonly in advanced disease as a stress response, but this is confounded by an increase in cortisol-binding globulin. Adrenocorticotropic hormone (ACTH) levels are usually normal or elevated in patients with elevated cortisol levels, although some reports of low levels exist. Some of these alterations may be mediated by multiple cytokines; both interleukin (IL)-1 and tumor necrosis factor (TNF) directly stimulate cortisol secretion, while IL-1 and IL-6 stimulate ACTH and corticotropin-releasing hormone (CRH) release. On the other hand, IL-2 and IL-4 increase sensitivity to cortisol. Increased cortisol levels may be a direct response to HIV infection itself. The HIV envelope protein gp120 increases cortisol. HIV Vpr, and Tat may increase cortisol sensitivity. An increased cortisol to dehydroepiandrosterone (DHEA) ratio correlates with body weight loss and malnutrition.

Subclinical abnormalities in hypothalamic pituitary axis dynamics are often present. Nearly all HIV-infected patients have a normal cortisol response to the standard dose (250 μg) ACTH stimulation testing. Other studies have shown that stimulation with low-dose cosyntropin (1 μg) yields results suggesting glucocorticoid insufficiency in roughly 10% to 20% of outpatients with HIV and up to 50% of critically ill HIV-infected patients, especially in HIV-infected patients with CD4 counts less than 500/μL. The abnormalities of decreased reserve either at the pituitary or at the adrenal level in HIV infection are similar to those that occur in other infections or serious illnesses.

Although clinically significant abnormalities in glucocorticoid secretion appear to be uncommon, subtle alterations in adrenal

biosynthesis may be more common. HIV-infected patients have reduced products of the 17-deoxysteroid pathway (corticosterone, deoxycorticosterone, and 18-hydroxydeoxycorticosterone) with normal or elevated products of the 17-hydroxy pathway (cortisol) before and after ACTH stimulation. It is not known whether this alteration represents an early indication of evolving adrenal insufficiency or is an adaptive response that shifts adrenal synthetic activity to steroids that are crucially needed under conditions such as HIV infection that impose physical stress. Twenty-four hour urine free cortisol levels do not predict the subtle adrenal alterations in HIV-infected patients, but may indicate adequate adrenal function.

Glucocorticoid resistance has been described in HIV-infected patients. This syndrome is characterized by symptoms of weakness, fatigue, weight loss, and hyperpigmentation, with elevated cortisol levels and mildly increased ACTH levels. Decreased lymphocyte glucocorticoid receptor affinity for glucocorticoids has been described in these patients. A similar state is seen in glucocorticoid-resistant asthma patients. Increased expression of the beta form of the glucocorticoid receptor, which inhibits the alpha form, has also been reported. Partial glucocorticoid resistance could explain the finding of increased basal cortisol in some HIV-infected patients, but the prevalence and clinical significance of this syndrome are uncertain.

Adrenal Androgens

HIV-infected patients have decreased basal adrenal androgen levels and impaired adrenal androgen responses to ACTH stimulation. Decreased adrenal androgens have been seen at all stages of HIV infection, as well as in HIV-negative intensive care unit patients. Thus, this change may not be specific to HIV infection but may instead be a feature of the physiologic response to illness. A fall in DHEA levels predicts progression to AIDS independent of CD4 cell counts. DHEA has been shown *in vitro* to inhibit HIV replication, which raised the possibility that the decreased DHEA levels observed in HIV-infected patients might influence the effects of the HIV infection. However, the efficacy of DHEA replacement in HIV-infected patients has never been demonstrated. Adrenal androgens may be decreased in women with HIV infection; low dose testosterone replacement has been shown to induce small improvements in lean mass, depression and sexual function.

Mineralocorticoids

Although electrolyte disturbances are not uncommon in HIV-infected patients, provocative testing of the mineralocorticoid axis has revealed few abnormalities. In contrast to glucocorticoid levels, basal and ACTH-stimulated aldosterone levels have been found to be normal in almost all HIV-infected patients studied, including inpatients. Case reports of hypo- and hyperaldosteronism appear in the literature, but there is no evidence that these disorders are increased in HIV infection. Longitudinal studies suggest that the aldosterone response to ACTH stimulation may diminish with progression to later stages of HIV infection in up to half of HIV-infected patients; however, basal levels of aldosterone and plasma renin activity remain normal, and clinically significant hypoaldosteronism does not develop.

TABLE 25–3 Usual pattern of adrenal hormones in HIV infection.

	Basal	After ACTH Stimulation
Glucocorticoids	Normal or ↑ cortisol; ↓ 17-deoxysteroids	Normal cortisol response
Mineralocorticoids	Normal	Normal
Androgens	↓	↓

↓, decreased; ↑, increased.

The usual pattern of alterations in adrenal hormones is shown in Table 25–3.

Medication Effects

Several medications used in the treatment of HIV-related disorders can alter glucocorticoid metabolism. Ketoconazole and itraconazole inhibit the cytochrome P450 enzymes P450scc and P450c11, decreasing cortisol synthesis and leading to adrenal insufficiency in patients with decreased adrenal reserve. Rifampin increases hepatic metabolism of steroids and may lead to adrenal insufficiency in patients with marginal adrenal reserve.

Megestrol acetate has intrinsic cortisol-like activity, decreasing serum cortisol and ACTH levels through suppression of the hypothalamic-pituitary-adrenal axis centrally. Patients taking megestrol have decreased cortisol and ACTH levels in response to metyrapone testing. They may become frankly Cushingoid as well and may show signs of adrenal insufficiency upon its rapid discontinuation. Some HIV protease inhibitors (PIs) block the metabolism of the inhaled steroids fluticasone and budesonide or injected triamcinolone by CYP3A4 and can lead to Cushing syndrome; there are reports of adrenal insufficiency after discontinuation of these therapies. PIs do not affect plasma cortisol levels. However, there is debate about whether these agents change urinary free cortisol or 17-hydroxycorticosteroid excretion.

Medications used to treat HIV-related illnesses can lead to electrolyte disturbances mimicking disorders of mineralocorticoids. Trimethoprim impairs sodium channels in the distal nephron, decreasing potassium secretion, which can result in hyperkalemia. Pentamidine has also been associated with hyperkalemia in rare instances, perhaps through nephrotoxicity. Sulfonamides are associated with interstitial nephritis and hyporeninemic hypoaldosteronism. Finally, amphotericin B causes renal potassium and magnesium wasting.

Medications used in HIV-infected patients that can affect the endocrine system are listed in Table 25–2.

Summary of Adrenal Disorders

In summary, there is little evidence for clinically significant impairment of adrenal steroid excretion in HIV infection. The subtle alterations in the glucocorticoid and androgen synthesis pathways may be an adaptive response to physiologic stress and may occur with other illnesses. Patients with HIV infection who

exhibit symptoms consistent with adrenal hormone deficiency should undergo provocative testing in the same manner as uninfected individuals. Patients with both low baseline and abnormal glucocorticoid or mineralocorticoid responses to provocative testing should be treated with physiologic replacement doses of oral glucocorticoids or mineralocorticoids. Patients on replacement should be covered with higher ("stress") doses of glucocorticoids during episodes of severe illness. Electrolyte abnormalities should prompt evaluation for medication effects and the presence of renal disease.

The HIV-infected patient with a minimally elevated or frankly high basal cortisol who does not increase significantly after ACTH stimulation poses a difficult problem. Most of these individuals have normal responses to prolonged ACTH stimulation, and seronegative patients with significant illness can have similar patterns that revert to normal after treatment of the illness. Chronic glucocorticoid therapy may have significant adverse consequences in these individuals, who are already immunocompromised. Most HIV-infected patients with this pattern and even many with baseline low levels during acute illness do not appear to require long-term glucocorticoid replacement. Consideration could be given to administering short courses of steroid therapy during significant illness for individuals with indeterminate stimulation results based on clinical judgment and individual patient circumstances.

BONE AND MINERAL DISORDERS

Osteopenia and Osteoporosis

There is evidence of low bone mass and increased osteoporosis in HIV-infected patients either without therapy or while receiving HAART, but the etiology remains controversial. The incidence of fractures may be increased, but there are too few studies of fracture incidence to know with certainty. Therapy with bisphosphonates is effective at increasing bone mineral density (BMD) in HIV-infected patients, but there are no convincing data on fracture prevention.

Prior to the introduction of HIV PIs, low BMD was not commonly observed in HIV-infected patients. The few clinical studies on the effects of HIV on bone mineralization suggested that HIV had relatively little effect on BMD. One study followed 45 HIV-infected men for 15 months and found no change in lumbar spine and hip BMD. However, histomorphometric analysis of bone biopsies from AIDS patients showed some evidence of decreased bone formation and bone turnover; these changes were more marked in more severely affected patients. HIV itself may have an indirect effect on BMD by activating the receptor activator of nuclear factor kappa B ligand (RANK-L), which activates osteoclasts and their precursors. Activated T cells express the receptor for RANK-L—that is RANK. The cytokines TNF-alpha and IL-1, whose levels may be increased in HIV infection, also have been shown to activate RANK-L.

After the introduction of PI-based HAART, some cross-sectional studies suggested that patients on PIs had a higher prevalence of osteoporosis and osteopenia compared to HIV-infected patients not taking PIs. However, other studies did not find a specific association of PIs with decreased BMD by dual energy x-ray absorptiometry (DXA). Most cross-sectional studies have found that both HIV-infected men and women on any form of HAART have lower BMD or a higher prevalence of osteopenia, compared to gender- and ethnicity-matched control subjects. In one meta-analysis of 20 papers, the prevalence of reduced BMD was 67%, and the prevalence of osteoporosis was 15% in HIV-infected patients, which represents a 6.4-fold greater prevalence of reduced BMD and 3.7-fold greater prevalence of osteoporotic T scores compared to controls. Furthermore, those treated with ARV had a 2.5-fold higher prevalence of reduced BMD than those who were ARV-naïve. A randomized longitudinal study showed that continuous ARV therapy, used to suppress HIV, was associated with greater loss of bone mass and perhaps more fractures than intermittent ARV therapy designed to reduce drug exposure. However, the absolute rate of BMD loss by DXA was actually quite small in those on continuous ARV therapy (−0.5%/year at the hip and −0.7%/year at the spine). Other data indicate that initiation of ARV use per se is associated with decreases in BMD. However, the role of specific ARV drugs as a cause of bone demineralization remains debated, as many studies fail to show an association between osteopenia and most drugs or drug classes. More recently, the use of tenofovir has been associated with greater loss of BMD during ARV therapy, with improvement on switch to tenofovir-sparing regimens. Tenofovir has also been shown to reduce BMD in HIV-seronegative persons when used as prophylaxis against HIV infection (PReP). Tenofovir can also induce hypophosphatemic osteomalacia.

A meta-analysis found that the effect of HIV infection was mostly accounted for by decreased weight. The increased smoking found among HIV-infected patients also contributes to low BMD. Other studies have found associations of hypogonadism, HCV coinfection, vitamin D deficiency, low albumin, and past steroid use with low BMD. Use of PI and inhaled or injected steroids with induction of Cushing syndrome has led to osteoporosis and fracture. In a large database analysis, the fracture rate was increased 1.7-fold in HIV-infected Caucasian men and 1.4-fold in HIV-infected Caucasian and African-American women compared to their respective controls. HCV coinfection is associated with increased fracture, even after adjusting for advanced liver disease and cirrhosis.

Treatment for HIV-associated osteoporosis with bisphosphonates has been examined. Multiple studies have shown that the bisphosphonates alendronate and zoledronic acid given, either with or without calcium and vitamin D supplementation, to HIV-infected patients with osteopenia or osteoporosis, improved BMD in the lumbar spine and hip after 1 to 2 years of therapy. Teriparatide also increased BMD in HIV infection.

The effects of treatments for other HIV-related problems on BMD or bone turnover markers have also been assessed. Eugonadal men with AIDS wasting syndrome treated with intramuscular testosterone for 3 months had small increases in lumbar spine BMD; however, high doses of testosterone were associated with decreases in high density lipoprotein (HDL) cholesterol levels. Recombinant growth hormone (GH) treatment for 24 weeks in HIV wasting syndrome did not affect BMD. Short-term growth

hormone-releasing hormone (GHRH), used to treat HIV-infected men with abdominal fat accumulation, was associated with an increase in bone turnover markers. The effect of GHRH use on BMD remains to be determined.

Osteonecrosis

Osteonecrosis has been reported both in adults and in children infected with HIV, and there is concern about the rising prevalence of osteonecrosis. The most likely reason for the apparent rise in prevalence of the syndrome is increased awareness among clinicians. The annual incidence has been reported to be between 0.08% and 1.3%, which is much higher than the annual incidence in the general population (0.01%-0.135%). A study of 339 HIV-infected patients examined with magnetic resonance imaging (MRI) found a prevalence of 4.4% for asymptomatic osteonecrosis.

The clinical presentation and anatomic distribution are similar to HIV-negative patients. The most common location for osteonecrosis is the hip, but other locations include the humeral heads, femoral condyles, and scaphoid and lunate bones. In HIV infection, the majority of large joint involvement is bilateral. Cases of osteonecrosis of the jaw/oral cavity have also been reported.

Plain film x-rays often underdiagnose osteonecrosis as changes are confined to advanced stage (III and IV) disease; MRI is more effective at finding early stage disease. Some data support treatment with bisphosphonates in early disease involving major joints to reduce pain, but there are inadequate randomized, controlled trials to assess whether such therapy prevents progression. Core decompression with or without bone graft has had some success. Total joint replacement may be the only recourse in late disease (eg, following collapse of the femoral head).

Possible etiologies include those that may be more frequent in patients with HIV infection (corticosteroid use, HIV-associated vasculitis, chemotherapy, and irradiation) and other traditional risk factors (ethanol use, sickle cell disease, hemoglobinopathies, and clotting disorders). An association between osteonecrosis and PIs was initially reported; however, subsequent studies have failed to show a causal relationship with PI use. Use of PI with inhaled or injected steroids inducing Cushing syndrome has led to osteonecrosis.

Calcium and Phosphate Homeostasis

Calcium, phosphate, and calciotropic hormone disturbances are associated with several HIV-related illnesses and medications. Hypercalcemia with an elevated 1,25-dihyroxyvitamin D level has been associated with both AIDS-related lymphoma and infections with *M. tuberculosis, M. avium-intracellulare, P. jiroveci, C. neoformans, and Coccidioides immitis.* This hypercalcemia is due to increased extra-renal 1-alpha hydroxylation of 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D by inflammatory macrophages or tumor cells. In CMV infection, activated T cells or proinflammatory cytokines may secrete factors that activate osteoclasts, causing hypercalcemia. The syndrome has also been reported after starting antiretroviral therapy, with improvements in lymphocyte function, as part of the immune reconstitution syndrome. Recombinant

human GH treatment has been associated with slight increases in serum calcium levels in some studies.

Mild hypocalcemia has been associated with HIV disease. HIV-infected patients have been shown to have low 1,25-hydroxyvitamin D and 25-hydroxyvitamin D levels, but normal vitamin D-binding protein levels, suggesting an impairment in 1-alpha hydroxylation and a compromised nutritional state. Low 1,25-hydroxyvitamin D levels are associated with advanced disease stage and increased TNF-alpha levels.

Inadequate parathyroid hormone (PTH) responses (ie, functional hypoparathyroidism) may also contribute to the hypocalcemia seen with HIV. PTH levels have been shown to be decreased in HIV-infected patients compared to controls. Parathyroid cells may be targeted by HIV infection: they express CD4 receptors which when activated affect PTH secretion. Invasion of the parathyroid gland by neoplasms and infections can also rarely occur.

Severe hypocalcemia is usually due to medications used to treat opportunistic infections. Variations on Fanconi syndrome are reported increasingly. Hypocalcemia and hypophosphatemia have been reported with tenofovir (used to treat HIV) and cidofovir (used to treat CMV). Hypocalcemia has also been seen during therapy with foscarnet for CMV retinitis. Foscarnet can complex ionized calcium and may have mineral wasting effects at the level of the renal tubule, leading to concurrent hypomagnesemia and hypokalemia. Hypocalcemia and hypomagnesemia have also been reported during pentamidine treatment, and the combination of foscarnet and pentamidine can result in severe, even fatal, hypocalcemia. Trimethoprim-sulfamethoxazole, ketoconazole, and aminoglycosides are also associated with hypocalcemia. Ketoconazole and rifampin can alter vitamin D metabolism but usually do not produce clinically significant effects. Ketoconazole can reduce serum levels of 1,25-dihydroxyvitamin D and lower total—but not ionized—calcium levels. Rifampin and rifabutin can decrease 25-hydroxyvitamin D levels, but usually do not appear to significantly change calcium or PTH levels; one case has been reported of severe osteomalacia attributed to rifabutin.

Medications used in HIV-infected patients that can affect the endocrine system are listed in Table 25–2.

GONADAL DISORDERS

Testicular Function

Testicular atrophy is common in AIDS patients. Histopathologic changes include decreased spermatogenesis, thickened basement membrane, and an interstitial infiltrate. Possible causes of testicular atrophy include opportunistic infection by *M. avium-intracellulare, Toxoplasma*, and CMV, hypogonadism, HIV infection itself, and chemotherapy toxicity. HIV-infected men generally have oligoteratospermia, and as the disease state becomes more severe, the sperm can become grossly abnormal. Hence, male fertility decreases with advancing disease.

Semen washing has been used as a technique for HIV-discordant couples to reduce the risk of viral transmission. There have been several thousand published cases of semen washing

followed by either intrauterine or *in vitro* fertilization. These observational reports have shown that semen washing is a relatively safe and effective means of establishing pregnancy in HIV-negative women. Still, there is concern that the risk of infection is not completely eliminated after washing.

Hypogonadism is a common finding in men, particularly with advanced HIV disease. Prior to the introduction of HAART, hypogonadism was observed in approximately 40% of HIV-infected men. With the widespread use of HAART, the prevalence of hypogonadism has declined to 20%. Early in HIV infection, testosterone levels are normal or even elevated. In the later stages of HIV infection, low serum testosterone levels are seen, often accompanied by symptoms of decreased libido or erectile dysfunction. Because hypogonadism is associated with loss of lean body tissue and muscle mass, it is thought to contribute to AIDS wasting syndrome. Hypogonadal men infected with HIV tend to have low or normal LH and follicle-stimulating hormone (FSH) levels, similar to that seen in hypogonadotropic hypogonadism. However, GnRH testing in hypogonadal men with HIV has been normal in most cases. Primary hypogonadism with elevated LH and FSH levels has been observed but is far less common.

Several etiologies of the hypogonadism in men infected with HIV have been proposed. In many cases, the cause of hypogonadism is multifactorial. Chronic illness and weight loss can alter the hypothalamic-pituitary axis. HIV infection increases the levels of the cytokines IL-1 and TNF, which have been shown to suppress hypothalamic pituitary function. Direct pituitary or hypothalamic invasion by toxoplasmosis or CMV is rare. There are divergent data on the extent to which treatment with ARV drugs increases testosterone levels.

Medications can also alter testicular function. Ketoconazole inhibits gonadal steroidogenesis, resulting in lower testosterone levels, oligospermia, and gynecomastia. Megestrol acetate, a progesterone-like agent, can lead to decreases in testosterone in HIV-infected men, perhaps through central feedback on gonadotropins. Opiate and methadone use in HIV may also cause low testosterone. Medications used in HIV-infected patients that can affect the endocrine system are listed in Table 25–2.

Most of the studies of testosterone replacement have been performed in the context of AIDS wasting syndrome in which muscle and fat mass are decreased. Testosterone therapy in hypogonadal men with HIV increases lean body mass, but only small improvements in strength are seen. Most studies find that testosterone replacement therapy results in improved sexual functioning, mood, and energy. As in the non-infected population, most HIV-infected patients who are given testosterone have not had their hypogonadism adequately evaluated.

Ovarian Function

Women represent one of the fastest growing populations with HIV infection. Two key issues are the effects of reproductive function on the natural history of HIV disease, and conversely, the effects of HIV on reproductive function. Recent studies found that women infected with HIV possess 30% to 50% lower HIV-1 RNA levels than men. Limited data suggests that the ovulatory

cycle affects HIV infection in women. Early studies suggested no association between the menstrual cycle and HIV-1 RNA levels but may have included women with anovulatory cycles. In ovulating women, HIV-1 RNA levels have been shown to decline from the early follicular phase to the luteal phase. Menopause does not seem to affect the progression of HIV disease or the response to therapy.

In addition to affecting HIV-1 RNA levels, the menstrual cycle may also affect the pharmacokinetics of drugs used to treat HIV infection. Zidovudine levels have been shown to be negatively correlated with estradiol levels, suggesting that the menstrual cycle may affect zidovudine glucuronidation. Further studies are warranted.

Pregnancy does not seem to have a deleterious effect on maternal progression of HIV disease. In those who do not have suppressed HIV, HIV-1 RNA levels may increase in the early postpartum period, and pregnancy may increase the rate of detection of HIV-1 in the genital tract. When HIV-1 is identified in the genital tract, the risk of perinatal transmission has been consistently shown to increase. Simple ARV regimens at the time of delivery have been shown to significantly reduce transmission of HIV to the child.

It is unclear whether HIV infection itself causes menstrual irregularities and/or amenorrhea. Early studies reported menstrual irregularities in HIV-infected women. However, these studies relied on self-report of menstrual symptoms and cycles, and many studies had enrolled an inadequate number of HIV-negative women as controls. Recent studies suggest that HIV may alter menstrual function only slightly. One large study of 802 HIV-infected and 273 HIV-negative women, matched for demographic characteristics, body mass index, and substance abuse, found that HIV infection increased the odds of having a very short or long cycle. There were no differences in amenorrhea between the two groups. Another study of 197 HIV-infected women and 189 HIV-negative women found that HIV was not associated with menstrual dysfunction.

It is also uncertain whether HIV has an effect on the onset and symptoms of menopause. The average age at menopause in HIV-infected women has been reported to be between 46 and 50 years. While some studies contrast that to reported average age at menopause in controls being slightly greater than 50, the range in multiple population studies is actually 48 to 51. The diagnosis of menopause is complicated by the occurrence of amenorrhea lasting more than 1 year (which meets the clinical definition of menopause) in HIV-infected women who do not have ovarian failure as indicated by normal FSH and LH levels. Factors associated with such amenorrhea include opiate use, low body mass index, and low serum albumin levels, suggesting hypothalamic dysfunction and malnutrition as causes rather than menopause. Other studies implicate increased smoking in HIV-infected persons and HCV coinfection as contributors to early menopause. There are conflicting data concerning whether HIV is associated with increased or decreased symptoms of menopause.

With HIV infection, fertility rates decline. Quantifying the rate of decrease in fertility in the United States has been difficult due to confounding factors such as sociodemographics, drug use,

weight loss, systemic illness, and sexually transmitted diseases. In Africa, where therapy for HIV was rare, there are numerous studies establishing an association between untreated HIV infection and infertility in virtually all age groups. In sub-Saharan Africa, HIV infection has been shown to account for a population-attributable decline in total fertility of 0.37% for each percentage point of HIV prevalence.

Like men, women infected with HIV have a decline in serum testosterone and adrenal androgen levels with progression of HIV disease and development of complications. Approximately 50% of women with AIDS-related wasting have low androgen levels. Androgen-deficient women may have subtle symptoms of decreased energy, libido, mood, strength, and bone mass. Clinical laboratory diagnosis of androgen deficiency remains difficult in the female population with advanced disease. Total testosterone levels may be increased due to elevations in sex hormone-binding globulin in HIV-infected women, whereas free testosterone levels for women are not well-standardized.

Hormonal contraception is affected by ARV therapy. The estrogen component is metabolized by CYP3A4 enzyme. Many of the PIs and some non-nucleoside reverse transcriptase inhibitors (NNRTIs) either inhibit or induce CYP3A4, thus altering ethinyl estradiol and norethindrone levels. Other medications used to treat opportunistic infections such as rifabutin and rifampicin can also induce CYP3A4. There are very few studies on the effects of these medications on ovulation in subjects taking oral contraceptives. The Centers for Disease Control and Prevention recommend using barrier methods or condoms in addition to or in place of oral contraception.

PITUITARY DISORDERS

Opportunistic Infections and Neoplasms

In autopsy series of AIDS patients, nearly 10% of pituitary glands demonstrate some degree of infarction or necrosis; infectious organisms such as CMV, *P. carinii, Cryptococcus, Toxoplasma*, and *Aspergillus* have been observed. Antemortem pituitary function was not reported in these patients.

Anterior Pituitary Function

Hypopituitarism appears to be rare in AIDS patients; stimulation with TRH, GnRH, or CRH results in normal pituitary responses in almost all patients. Likewise, prolactin levels generally have been found to be normal, with a normal response to TRH stimulation.

The GH axis has received particular attention in children with HIV who can have poor growth velocity, especially when symptomatically ill. Most of these children demonstrate normal GH levels, although low insulin-like growth factor (IGF)-I levels can be seen. Low IGF-I in the setting of relatively normal GH is often seen in states of malnutrition, which may partially explain this finding. In adults, IGF-I levels may be low in malnourished or symptomatic HIV-infected patients but are usually normal in clinically stable individuals. Circadian GH secretion does not

appear to be altered in adults with HIV infection, except in the presence of visceral obesity.

Posterior Pituitary Function

Posterior pituitary function may be altered in HIV infection. Hyponatremia is common in both inpatients and outpatients with HIV. Inappropriately high serum antidiuretic hormone levels have been seen in euvolemic AIDS patients with hyponatremia; however, many of these patients had pulmonary or cerebral infections, which themselves can cause the syndrome of inappropriate antidiuretic hormone secretion. As with any patient, destruction of the neurons with projections to the posterior pituitary gland by infection or tumor can lead to central diabetes insipidus; this appears to be rare in HIV infection.

AIDS WASTING SYNDROME

Prior to the introduction of HAART, weight loss and tissue wasting were commonly observed complications of AIDS in up to 30% of patients infected with HIV. A characteristic feature of this syndrome is wasting of muscle, which is directly associated with increased morbidity and mortality. Muscle is disproportionately lost compared to fat in men. In contrast, women start with higher fat mass, and fat depletion dominates muscle loss until fat reserves are decreased. Weight loss can either be gradual or rapid in nature. Gradual weight loss has been associated with gastrointestinal disease, whereas rapid weight loss is usually seen during opportunistic infections.

The pathogenesis of wasting remains poorly understood. Several etiologies have been proposed including starvation, cachexia, hypermetabolic state, and hypogonadism. Starvation implies solely a limitation of calories, which may occur due to decreased appetite or decreased intake due to gastrointestinal infections with marked diarrhea. Cachexia refers to a state in which muscle is disproportionately lost due to direct effects on muscle. A hypermetabolic state has been proposed as a cause of AIDS wasting syndrome, because resting energy expenditure (REE) has been observed to be increased in HIV infection. However, the increase in REE is also seen in HIV-infected subjects with stable weight. HIV patients with either opportunistic infections or gastrointestinal disease have decreased caloric intake, but do not show the protective decrease in REE that occurs in pure starvation. Hence, the combination of decreased caloric intake and hypermetabolism is the driving force for wasting. The persistent metabolic defects make recovery from wasting incomplete, and recurrent events lead to progressive, step-wise wasting.

Hypogonadism is a prominent finding in up to 50% of men and women with AIDS wasting. A direct correlation between testosterone levels and decreased muscle and fat mass exists. Resistance to GH has also been demonstrated in patients with more than a 10% loss in body weight.

Because a 10% loss of weight is associated with increased morbidity and mortality in patients infected with HIV, clinical trials have targeted the restoration of appetite, weight, and lean body mass. The cannabinol derivative dronabinol increased subjective appetite in patients with AIDS wasting syndrome, but had little or

no effect to promote weight gain. Megestrol acetate increased appetite and weight, but resulted in mostly gain of fat.

Currently, recombinant GH therapy is the only Food and Drug Administration-approved treatment specifically approved for AIDS wasting syndrome that increases lean body mass. At high doses (0.1 mg/kg/d), 3 months of GH therapy increased lean body mass by 3 kg and weight by 1 kg. However, patients did experience side-effects including swelling and joint pain. In a lower dose study (1.4 mg/d), patients had no significant increase in lean body mass after 12 weeks, which suggests that high-dose GH therapy is required. Therapy with recombinant GH is costly, especially if given at higher doses.

Androgen therapy has been studied extensively, because almost half of patients with AIDS wasting syndrome exhibit decreased levels of testosterone. In a meta-analysis of six clinical trials, testosterone therapy was shown to increase lean body mass and weight, especially when the therapy was given intramuscularly as opposed to transdermally. However, there have been no studies in which subjects have been randomized to different routes of administration of testosterone. Because testosterone may decrease HDL cholesterol levels, there is some concern about the long-term effects of testosterone therapy on lipid profiles and cardiovascular risk. However, no other side-effects were reported in these small, short-term studies.

Anabolic steroids have similar effects on the AIDS wasting syndrome. Nandrolone decanoate is an injectable testosterone derivative that is approved for treating anemia in patients with chronic renal failure. Studies of nandrolone decanoate in men found that weight, lean body mass, and strength significantly increased without evidence of acute toxicity. Therapy using oral anabolic steroids oxandrolone and oxymetholone also increases lean body mass and weight in men and women, but it is associated with significant hepatic transaminitis as well as dyslipidemia, especially low HDL.

Another approach has been to use anticytokine therapies to treat patients with AIDS wasting syndrome, but the results have been disappointing. Two studies reported small increases in weight after treatment with thalidomide, a TNF inhibitor. However, TNF levels actually increased and the patients experienced significant side effects including rash, fever, and increased viral load.

ABNORMALITIES OF FAT DISTRIBUTION ASSOCIATED WITH HIV

Shortly after the introduction of HAART with PIs, a variety of symptoms appeared including abdominal obesity, lipoatrophy in the extremities and face, dorsocervical fat pads (buffalo hump), gynecomastia, breast enlargement, increased neck circumference, and lipomas along with insulin resistance and dyslipidemia. These changes were proposed to be a single syndrome reminiscent of the metabolic syndrome and named HIV lipodystrophy.

In retrospect, some of these changes are not related to HAART with PIs. The appearance of lipoatrophy, fat accumulation in the dorsocervical fat pad, lipomas, and breast hypertrophy was reported in patients on two NRTIs before the introduction of HAART or PIs. The presence of dorsocervical fat pads and lipomas is striking, but their prevalence is low.

Furthermore, the original description of the HIV lipodystrophy syndrome has been refuted, as it has been shown definitively that the major changes in fat (lipoatrophy and abdominal obesity) are not part of the same syndrome, but have different causes. Unfortunately, there is still no widely-accepted definition of HIV lipodystrophy that is used in the literature. Depending on the definition of HIV lipodystrophy, the prevalence of these findings varies widely (11%-83%). Part of the confusion is due to differences in methodology used to report fat loss or gain. Many early studies relied on the patient's self-report and/or the clinician's subjective assessment. Later studies quantified fat by anthropometrics or DXA. Neither technique permits the isolation of subcutaneous adipose tissue (SAT) in the abdomen and trunk from visceral fat (VAT); therefore computed tomography and MRI are needed and have been utilized more recently to separately quantify fat in these compartments. Because changes in fat may take years to develop, many studies in the field are cross-sectional rather than prospective in design. In these cross-sectional studies, it is unclear how to match patients to a control group with "normal" fat distribution, given that most healthy adults are overweight.

Despite these challenges, it has become apparent that fat loss with the lower body affected more than the upper body, but involving the face, does occur with increased frequency in patients taking HAART. In one study, 50% of HIV-infected subjects fell below the lowest decile (10%) of leg SAT in controls. In contrast, it is not clear that VAT mass is increased in HIV infection. Furthermore, recent cross-sectional and prospective data showed that increased VAT is not linked to loss of SAT.

Several etiologies have been proposed for the development of lipoatrophy associated with HIV, but to date no mechanism has been definitively linked to the disorder. The thymidine-based nucleoside analogs, stavudine and zidovudine, have the strongest association with the development of lipoatrophy. These NRTIs have been proposed to cause lipoatrophy through mitochondrial injury by inhibiting mitochondrial DNA polymerase-γ or depleting mitochondrial DNA. Fortunately, the use of stavudine and zidovudine has decreased. Abacavir is the NRTI with the least effect on fat. Some PIs (eg, indinavir) and NNRTIs (eg, efavirenz) may contribute secondarily to the lipoatrophy observed with NRTI therapy, but there are not enough trials to understand fully the effects of each ARV drug.

The development of central lipohypertrophy is even less well understood. While central lipohypertrophy has been attributed to PI use, the evidence is controversial. In the first reports of appearance of buffalo hump, half of the patients were not on PI therapy or HAART. Recent data suggest that increased VAT may be due to restoration to health and normal aging.

Much research has focused on the treatment of the fat changes associated with HIV. Fat loss, especially in the face, can reveal the HIV status of a patient and lead to social stigmatization. The metabolic sequelae can also predispose the patient to the development of hypertriglyceridemia and atherosclerosis. One proposed strategy is to reverse the peripheral lipoatrophy by switching ARV drugs. Substituting another NRTI for a thymidine NRTI, such as

stavudine or zidovudine, causes a minimal, albeit statistically significant increase in subcutaneous fat. In one study, switching from stavudine or zidovudine to abacavir caused a 0.4-kg increase in total limb fat after 24 weeks and a 1.3-kg increase in total limb fat after 104 weeks. However, that gain is insufficient to reach the levels seen in control populations. Lipoatrophy does not fully resolve. Switching from PIs to nevirapine, efavirenz, or abacavir did not improve lipoatrophy. The newer NRTIs, tenofovir and emtricitabine, have been associated with less peripheral lipoatrophy than stavudine.

Thiazolidinediones are potentially attractive agents for treatment of lipoatrophy and insulin resistance. As ligands for the transcription factor peroxisome proliferator activator receptor gamma (PPARγ), thiazolidinediones promote adipocyte differentiation. Unfortunately, two randomized studies found no improvement and a third showed minimal improvement in lipoatrophy after thiazolidine treatment for at least 24 weeks. However, the studies did show the expected improvement in insulin resistance.

Reconstructive therapy is available for peripheral lipoatrophy. Injectable skin fillers for the face include bioabsorbable and permanent fillers. Bioabsorbable fillers such as poly-L-acetic acid stimulate collagen production. In one open-label, single-arm study, 43% of patients had an increase in cutaneous thickness greater than 10 mm in the face after 96 weeks of therapy. The effects on collagen production last for approximately 18 months. Hyaluronic acid has also been used to a lesser extent in the treatment of facial lipoatrophy. Side-effects include easy bruisability, redness, and swelling. Permanent fillers are generally not recommended due to the dynamic nature of fat loss in peripheral lipoatrophy. More invasive procedures include autologous fat transplants.

Central lipohypertrophy has been treated with GH and GHRH analogues. At high doses (6 mg/d), GH decreased VAT by 21% to 24% but also decreased SAT. However, a significant number of patients experienced joint swelling, fluid retention, and deterioration in glucose tolerance. A lower dose of GH (3 mg/d) also reduced VAT and SAT, although to a lesser extent. With further dose reduction, GH (1 mg/d) decreased SAT, but did not significantly alter VAT content. A GHRH analogue, tesamorelin, limits the side-effects of excessive GH by increasing the endogenous pulsatility of GH secretion. Tesamorelin (2 mg/d) given for 26 weeks decreased VAT by 15%, but had little effect on SAT. Unlike GH therapy, tesamorelin did not induce insulin resistance; there were fewer GH-like symptoms of joint and fluid retention. GH or tesamorelin therapy is very expensive, and the effects reverse when therapy is stopped.

DISORDERS OF GLUCOSE AND LIPID METABOLISM

Shortly after the introduction of HAART, reports of impaired fasting glucose, glucose intolerance, diabetes, and hyperlipidemia appeared in the HIV literature. Given the close temporal relationship to the introduction of PIs, studies focused on the association with PIs. However, other factors such as restoration of health, genetic predisposition, immune reconstitution, and body composition changes (including lipoatrophy and VAT hypertrophy) may also contribute to the disturbances in metabolism. To clarify these issues, several groups have given HIV antiretroviral drugs to healthy, HIV-seronegative volunteers in order to define direct drug effects. Another approach has been "switch studies," where patients with suppressed viral load are randomized to change antiretroviral drugs, eliminating the effects of suppression of HIV. In the sections below, the results in HIV-infected subjects will be compared to those in healthy volunteers to understand the effects of drug versus disease. These studies have led to the conclusion that metabolic effects are specific to certain drugs and not an effect of the PI class as a whole.

Insulin Resistance, Glucose Intolerance, and Diabetes

With the reports of rapid onset of diabetes after introduction of HAART with a PI, researchers looked for effects of PI on insulin resistance, hepatic glucose production, and insulin secretion.

Several factors contribute to the development of insulin resistance in the setting of HIV. Unlike other infectious states, in which insulin resistance is common, early in the epidemic, AIDS was found to be associated with an increase in insulin sensitivity (Table 25–4). Compared with healthy controls, insulin sensitivity was higher in stable patients with symptomatic HIV. However, with asymptomatic HIV infection, another early study found that there was no change in insulin sensitivity compared with healthy controls. Insulin resistance is common in healthy subjects. Thus, improvement of HIV infection alone may contribute to an observed decrease in insulin sensitivity. Subsequent studies found insulin resistance in ARV-naïve, HIV-infected subjects. Insulin resistance correlates with CD4 count.

Body composition influences insulin sensitivity. Patients studied early in the epidemic, were thin, if not cachectic. In recent studies body weight was higher, even in many ARV-naïve patients. Increased VAT has been linked to insulin resistance and worsened glucose tolerance in subjects with or without HIV disease. Upper

TABLE 25–4 Effect of HIV and AIDS status on glucose and lipid metabolism prior to the introduction of HAART.

	Insulin Resistance	Total Cholesterol	Triglyceride	VLDL	LDL	HDL
HIV	↔	↓	↑ (8%)	↑ (7%)	↓ (16%)	↓↓↓ (36%)
AIDS	↓	↓	↑↑↑↑ (99%)	↑↑↑↑ (98%)	↓↓↓ (31%)	↓↓↓ (37%)

↔, No effect; ↑ or ↓, 0%-20%; ↓↓↓, 30%-40%; ↑↑↑↑, >40%. Percentage increases or decreases are denoted when available in the literature.

trunk and neck SAT are also independently, strongly associated with insulin resistance. Likewise, severe lipoatrophy has been linked to insulin resistance and glucose intolerance regardless of HIV status. The lesser levels of lipoatrophy seen in most HIV-infected patients also contribute. Other traditional risk factors, such as physical inactivity, also play a role in HIV-infected subjects. It should be recognized that each of these factors may contribute in an additive way to insulin resistance.

Much attention has focused on the role of individual therapies in the induction of insulin resistance. Some PIs have been reported to decrease insulin-mediated glucose disposal (M/I during the hyperinsulinemic, euglycemic clamp, a technique during which insulin is infused at a steady rate and glucose infused to maintain euglycemia, which directly measures insulin action). In a double-blind, placebo-controlled study in healthy normal volunteers, a single dose of indinavir has been shown to decrease insulin-mediated glucose disposal by 34% (Table 25–5). Indinavir for 4 weeks has also been shown to cause a 17% decrease in insulin-mediated glucose disposal as well as deterioration in glucose tolerance. A single, full dose of ritonavir decreased insulin sensitivity by 15%. Lopinavir boosted by lower dose ritonavir likely has less of an effect on insulin sensitivity (see Table 25–5). In two studies, lopinavir/ritonavir given for 4 weeks caused no change in insulin sensitivity, whereas in shorter studies, lopinavir/ritonavir given for 1 to 5 days was associated with a 13% to 24% decrease in insulin sensitivity.

Of note, not all PIs decrease insulin-mediated glucose disposal. In double-blind, placebo-controlled studies, atazanavir and amprenavir had no effect on insulin sensitivity in healthy normal volunteers (see Table 25–5). NNRTIs have not been associated with insulin resistance. In studies where PIs are replaced with NNRTIs, insulin resistance improved.

The mechanism of insulin resistance with PIs includes the acute blockade of the peripheral insulin-regulated glucose transporter (GLUT)4. In vitro studies have shown that PIs (indinavir, ritonavir, and amprenavir) selectively inhibit 2-deoxyglucose transport into 3T3-L1 adipocytes without affecting early insulin signaling events or the translocation of intracellular GLUT4 to the surface. Indinavir has also been shown to block partially GLUT2, the glucose transporter postulated to be involved in glucose-sensing in the pancreas and regulation of insulin secretion. Recently, an analog of the peptidomimetic phenylalanine moiety found in all PIs has been shown to inhibit GLUT4-induced glucose transport in vitro. Because serum levels of PIs vary, some PIs, such as amprenavir, may block GLUT4 in vitro but have no effect in patients.

Increased insulin resistance has also been found in HIV-infected subjects on NRTI therapy. However, it is unclear if the effects of NRTIs are a direct effect of the drug, reactivation of the immune system, restoration of health, or changes in body composition. When stavudine was given to healthy volunteers, there was no decrease in M/I.

In addition to peripheral insulin resistance, impairment in insulin secretion was reported in HIV patients on PI therapy. In HIV-infected subjects treated with several PIs, beta cell function assessed by first-phase insulin secretion showed a 25% decrease.

However, insulin secretion in the HIV-infected patients was higher than controls before PI therapy and was reduced only to that of controls after PI therapy. The HIV-infected patients had suppression of HIV RNA levels and increases in CD4 counts. More recently, healthy normal volunteers were given lopinavir/ritonavir for 4 weeks, and no effect was seen on first-phase insulin secretion. Thus, it is unlikely that currently used PIs alter insulin secretion.

Hepatic glucose production is also increased in some patients on PIs. Endogenous glucose production, comprised mostly of hepatic gluconeogenesis and glycogenolysis, is the largest determinant of fasting glucose. In studies of healthy normal volunteers, indinavir increased endogenous glucose production in the fasting state and blunted insulin suppression of endogenous glucose production during a hyperinsulinemic, euglycemic clamp. In humans, full dose ritonavir has a small detrimental effect on endogenous glucose production, while amprenavir had no effect.

Increased plasma levels of the NNRTI efavirenz, particularly in those with the slow metabolizing mutation in CYP2b6 are associated with higher glucose levels. In contrast, there was no relation between levels of lopinavir or stavudine and glucose parameters.

Adipocytokine levels in HIV infection may explain some of the results in glucose metabolism. Adiponectin, a hormone secreted by adipocytes, has been shown to increase peripheral and hepatic insulin sensitivity. Adiponectin levels inversely correlate with the amount of VAT in HIV-negative subjects. In patients with HIV-associated lipoatrophy, adiponectin levels are reduced; thus, low adiponectin levels have been proposed to mediate some of the insulin resistance in HIV. The mechanism which reduces adiponectin levels is unknown. Some have attributed the reduction to PI therapy. In vitro studies of cultured fat cells have suggested that PI treatment suppresses adiponectin mRNA and protein expression. However, two studies in healthy normal volunteers found that in fact adiponectin levels were increased during chronic treatment with the PIs indinavir or lopinavir/ritonavir; because adiponectin increases insulin sensitivity, the higher levels may explain why less insulin resistance is seen after 4 weeks of treatment compared to acute dosing. Levels of leptin, another hormone secreted by adipocytes, correlate with insulin resistance. However, some diseases with very low leptin levels also have insulin resistance. Leptin levels have been shown to be decreased in HIV patients with peripheral lipoatrophy.

HCV and possibly Hepatitis B (HBV) coinfection are also associated with increased fasting glucose levels and prevalence of diabetes. Treatment of HCV mono-infection with newer antiviral drugs reduces fasting glucose levels.

Epidemiological studies of diabetes in HIV infection have shown contradictory results for specific antiretroviral drugs, but similar results for other risk factors. Some studies have implicated the PIs indinavir, ritonavir or saquinavir, and the NRTIs stavudine or didanosine as associated with diabetes, but individual studies have found different drugs to be associated with diabetes and some found the very same drugs to be protective. Pre-therapy BMI, use of effective combination therapy, increase in or higher CD4 counts, weight gain with therapy, decreased physical activity, the presence of lipoatrophy, higher triglycerides and older age are associated with incident diabetes.

TABLE 25–5 Effect of protease inhibitors (PI), nucleoside reverse transcriptase inhibitors (NRTI), non-nucleoside reverse transcriptase inhibitors (NNRTI), and integrase inhibitors on glucose and lipid metabolism.[1]

Class/Drug	Fasting Glucose	Insulin[2] Resistance		TG	LDL	HDL
Protease Inhibitors						
Indinavir	↑ (HIV-negative)	↑↑↑ (−34% M/I) (HIV-negative)		↔ (HIV-negative)	↔ (HIV-negative) ↑ (HIV-infected)[3]	↔↑ (0%-20%)
Ritonavir	↔	↑↑ (−18% M/I) (HIV-negative)		↑↑↑ (150% HIV-negative)	↔(HIV-negative) ↑(HIV-infected)[3]	↓ (0%-5%)
Lopinavir/ ritonavir	↔ (HIV-negative)	↔/↑[4] (−13%-24% M/I[4]) (HIV-negative)		↑↑ (80% HIV-negative)	↔(HIV-negative) ↑(HIV-infected)[3]	↔
Tipranavir/ ritonavir	↔	NA		↑↑ (80% HIV-infected)	NA	NA
Atazanavir/ ritonavir	↔ (HIV-negative)	↔ (HIV-negative)		↑	↔↑	↔↑ (0%-20%)
Darunavir/ ritonavir	NA	↔/↑		↑	↔↑	NA
Amprenavir	↔	↔ (HIV-negative)		↔↑	↑ (HIV-infected)	↑ (0%-20%)
Nelfinavir	↔	↔		↔↑	↑	↑ (2%-20%)
Saquinavir	NA	NA		NA	NA	NA
Atazanavir	↔	↔ (HIV-negative)		↔ (HIV-negative)	↔(HIV-negative) ↔(HIV-infected)	↔/↑ (0%-20%)
NNRTI						
Efavirenz	↔/↑	NA	NA	↔/↑↑	↔/↑	↑↑ (15%-23%)
Nevirapine	↔	↔		↔	↔/↑	↑↑↑ (49%)
Etravirine				↔	↔/↑	↑
Rilpivarine				↔	↔/↑	↑
NRTI						
Stavudine	NA	NA		↔/↑	NA	NA
Tenofovir	↔	NA		↓	↓	↓
Abacavir	↔	↔		↔/↑	↔/↑	
Integrase Inhibitors						
Raltegravir	↔	↔		↔	↔/↑	↔/↑
Dolutegravir	NA	NA		↔/↑	↔/↑	↔/↑
Elvitegravir	NA	NA		↔	NA	NA

[1]Percentage changes are provided when available from the published literature. The most effective measure of a drug's direct effects comes from studies where individual drugs were given to healthy, HIV-negative volunteers. Some effects were not precisely quantified in publications or have to be inferred from comparative regimens or switch studies. ↑ or ↓, small; ↑↑ or ↓↓, moderate; ↑↑↑ or ↓↓↓, large change. ↔/↑ signifies disagreement in the literature.

[2]Insulin resistance is expressed as M/I from euglycemic hyperinsulinemic clamp data as explained in the text. An increase in insulin resistance is reflected in a negative M/I and an upward arrow.

[3]LDL increases in patients with HIV infection who start with low LDL. There is little effect of PI in healthy, HIV-negative subjects.

[4]In healthy, HIV-negative subjects given one to five doses. No effect after 4 weeks.

Abbreviations: HDL, high-density lipoprotein; LDL, low-density lipoprotein; NA, not available; TG, triglyceride.

Current guidelines set forth by the Infectious Disease Society of America and similar European guidelines suggest measuring fasting glucose levels and/or hemoglobin A1c before and during anti-retroviral therapy. Given the controversies over which ARV induce diabetes, current guidelines refer to all patients with HIV. Hemoglobin A1c may not accurately reflect glucose levels in the presence of anemia of HIV.

Treatment of diabetes in HIV infection should follow the American Diabetes Association guidelines. There are studies of

therapy of patients who have HIV-associated lipoatrophy and lipohypertrophy. Thiazolidinediones improve insulin resistance in patients with HIV-associated lipoatrophy and lipohypertrophy. Proliferation of lipomas has been reported in one patient with HIV-associated lipoatrophy. Given the recent findings that thiazolidinediones decrease BMD and may increase fracture risk, caution is warranted in HIV-infected patients as they may be at higher risk for bone loss and fracture. Metformin decreases hepatic glucose production and peripheral insulin resistance. Metformin should be used with caution in combination with NRTI therapy, because there is an increased rate of lactic acid production with NRTIs, especially stavudine and didanosine. Cases of severe acidosis have been reported when these drugs were used in combination with metformin. Dolutegravir, an HIV integrase inhibitor, slows metabolism of metformin by 50%; lower doses of metformin should be used in patients on dolutegravir. Leptin has been shown to improve glucose tolerance in lipoatrophic patients with or without HIV, but is not yet approved by the FDA.

Medications used to treat opportunistic infections are associated with hyperglycemia and hypoglycemia. Pentamidine, including that administered by aerosol delivery systems, causes pancreatic beta cell toxicity, acutely leading to hypoglycemia. Over the long term this medication causes diabetes mellitus. Hypoglycemia during pentamidine treatment is associated with increased length of treatment, higher cumulative doses, and renal insufficiency. Patients who develop hypoglycemia on pentamidine are at increased long-term risk of developing diabetes mellitus. These patients have low C peptide levels, suggesting beta cell destruction. Pentamidine, trimethoprim-sulfamethoxazole, and the nucleoside analogs didanosine and zalcitabine have been associated with acute pancreatitis. Megestrol acetate, which has intrinsic glucocorticoid activity, may be associated with diabetes mellitus in HIV-infected patients, perhaps through its glucocorticoid activity, although the rate of hyperglycemia in controlled clinical trials appears to be low. GH can also cause insulin resistance, leading to hyperglycemia and diabetes. Medications used in HIV-infected patients that can affect the endocrine system are listed in Table 25–2.

Lipid Disorders

Alterations in lipid and lipoprotein profiles are common in patients infected with HIV. The observed changes can be due to HIV infection itself, antiviral medications, body composition changes, and immune reconstitution. A rational approach to disturbances in lipid metabolism is to assess each of the factors in a given patient. The following section reviews the lipid and lipoprotein profiles individually, with an emphasis on studies prospectively measuring fasting lipid levels.

HIV infection is associated with a mild increase in triglyceride and very low density (VLDL) cholesterol levels (see Table 25–4). Triglyceride and VLDL cholesterol levels rise in association with advancing HIV disease and correlate with HIV RNA levels. The host response to viral infections like HIV is induction of interferon alpha; in those with detectable viral loads, interferon alpha levels correlate with triglyceride levels. Decreased clearance of triglycerides and, to a lesser extent, increased VLDL production. Interferon alpha has been shown to induce these changes in animals.

Several antiviral medications can increase triglyceride levels. Full-dose ritonavir can cause a two- to threefold increase in triglyceride levels, probably by increased VLDL production (see Table 25–5). Because of its ability to inhibit the hepatic enzyme CYP3A4, ritonavir is used to increase the pharmacologic doses of other PIs metabolized through the same cytochrome system. Boosting doses of ritonavir (100 mg twice daily) have also been shown to increase triglycerides, albeit to a lesser extent. The combination lopinavir/ritonavir given to healthy normal volunteers increased triglycerides and VLDL cholesterol levels by 83% and 33%, respectively. Ritonavir-boosted tipranavir and fosamprenavir produce similar increases to those of lopinavir/ritonavir. Ritonavir-boosted atazanavir and darunavir appear to induce less increase in triglycerides. Not all unboosted PIs alter triglyceride levels; in healthy normal volunteers, administration of indinavir and atazanavir resulted in no change in triglyceride levels (see Table 25–5). The data are not clear for unboosted amprenavir and nelfinavir, but they are not commonly used as monotherapy.

The NNRTI efavirenz is associated with increased triglycerides. Again, this is not a class effect as other NNRTIs (eg, nevirapine, rilpivirine, and etravirine) have not been associated with alterations in triglyceride metabolism. The HIV integrase inhibitors (eg, raltegravir, dolutegravir, and elvitegravir) have no effect on triglycerides.

The effects of NRTIs on lipid metabolism have not been well studied. In some, but not all studies, stavudine use was associated with increased triglyceride levels. The most informative trial compared stavudine with tenofovir and found an increase in triglycerides in the stavudine arm, but not in the tenofovir arm. Given that all subjects got efavirenz, a likely interpretation is that there was a lipid-lowering effect of tenofovir. Other data support this interpretation, as switch studies or adding tenofovir to an effective ARV regimen lower lipids, but the effect on triglycerides in those studies is small.

The original clinical syndrome of HIV lipodystrophy was reported to be associated with increased triglyceride levels. One study found that 57% of patients with *both* peripheral lipoatrophy and central lipohypertrophy had triglyceride levels above 300 mg/dL. It has long been recognized that visceral obesity in the general population is associated with high triglycerides. However, recent data suggest that lower body fat is protective, as increased levels of lower body fat are associated with lower triglycerides in both HIV-infected patients and controls. Therefore, the loss of lower body fat in HIV-associated lipoatrophy is another reason why triglycerides are high in HIV infection. Hypertriglyceridemia is well known to be multifactorial; genes, diet, alcohol, and physical activity play a role. In HIV infection, one can add the synergistic effects of the host response to HIV itself, treatment with ritonavir and efavirenz, and HIV-lipoatrophy to the pathogenesis.

After the introduction of HAART, a reported increase in low density lipoprotein (LDL) cholesterol levels was largely attributed to PI therapy. It is now clear that factors other than PIs contribute to this rise in LDL cholesterol levels. In the early stages of HIV infection, LDL cholesterol levels fall (see Table 25–4). With effective therapy, LDL levels rise in response to suppression of HIV, independent of the type of therapy. Most, but not all PIs have

been associated with increases in LDL levels, but average levels are not high. Studies in healthy normal volunteers have provided insight into the direct effects of PIs on cholesterol metabolism apart from those associated with HIV infection. Indinavir, ritonavir, lopinavir/ritonavir, and atazanavir all have no effects on LDL levels in healthy normal volunteers (see Table 25–5). In patients with HIV infection, treatment with the NNRTI nevirapine raises LDL levels. Studies that involve switching patients from PIs to the NNRTIs nevirapine or efavirenz found that LDL levels do not change. Hence, the increase in LDL seen in HIV infected patients is not solely an effect of PI therapy, but likely represents suppression of HIV and restoration to health. However, in switch studies, some drugs have been shown to have less of an effect on LDL (the PIs atazanavir, saquinavir, and darunavir; the NNRTIs rilpivirine and etravirine; and the integrase inhibitors). Tenofovir lowers LDL levels in both HIV-infected and non-infected subjects.

HIV infection and the NNRTIs have significant effects on HDL metabolism. Early in the course of HIV infection, prior to the appearance of clinically evident disease, HDL cholesterol levels decline to levels around 25 to 35 mg/dL (see Table 25–4). With advancing HIV disease, HDL levels continue to decline to less than 50% of baseline values. The pathogenesis of these changes is not well understood. Shortly after the introduction of PIs, one cross-sectional study reported decreased HDL levels in HIV-infected patients. However, subsequent studies have failed to show a decrease in HDL levels in either HIV-infected patients or healthy normal volunteers. Indeed, some studies have found modest increases (13%-21%) in HDL levels during treatment with indinavir, nelfinavir, amprenavir, and atazanavir therapy in prospective studies of HIV-infected patients.

More impressive is the nearly 50% increase in HDL cholesterol seen during treatment with the NNRTI nevirapine. Efavirenz has also been shown to increase HDL levels by 15% to 23%. When patients were switched from PIs to the NNRTIs efavirenz and nevirapine, somewhat smaller increases were seen again supporting the concept that HAART with PI induces small increases in HDL. The story is not as clear for the NNRTIs rilpivirine and etravirine, as HDL levels decrease when stable patients are switched from efavirenz to those drugs.

The integrase inhibitors raise HDL, but less than the NNRTI and possibly less than PI. Tenofovir is the NRTI with the least effect on HDL. Combination therapy with tenofovir as the backbone raises HDL less than other NRTI. Switching to tenofovir results in small decreases in HDL.

In some studies, HDL levels have been reported to be lower in patients with HIV-associated lipodystrophy, but the extent to which that decrease is due to HIV or the effects of ARV are not clear. Visceral obesity and upper trunk fat are associated with lower HDL levels.

Several changes in lipoprotein structure and function occur in HIV infection that might promote atherosclerosis. The hypertriglyceridemia of AIDS is accompanied by increases in the prevalence of small dense LDL or LDL-B, even in the absence of obesity. LDL-B is also more common with ARV regimens that increase triglycerides.

Levels of oxidized LDL, which promotes atherosclerosis more than native LDL, and a marker of LDL oxidation, IgG against oxidized LDL, are increased in HIV infection. Their levels are higher on PI regimens compared to NNRTI regimens. HDL from those with HIV infection has less paraoxonase-1 (PON1) activity, which protects LDL from oxidation. PON1 activity has been shown to be proportional to CD4 count and inversely proportional to HIV viral load. Treatment may not restore PON1 to normal, although the increase is highest with NNRTI. PON-3, which has less of a protective effect, is increased in HIV and restored toward normal by NNRTI. Myeloperoxidase levels are elevated; the extent to which they improve with ARV is not clear. Lipoprotein associated phospholipase A2 (Lp-PLA2 or platelet activating factor acetyl-hydrolase) activity is increased in HIV infection. When lipoproteins are oxidized, Lp-PLA2 produces lysophosphatidylcholine, which is pro-atherogenic. Lysophosphatidylcholine levels are elevated in HIV. PI- and NNRTI-based therapy does not restore Lp-PLA2 activity to normal, but integrase inhibitors have a greater effect.

Most studies find that HDL from untreated HIV-infected patients has decreased ability to mediate cholesterol efflux from macrophages, the first step in reverse cholesterol transport, which protects against atherosclerosis. All studies show that antiretroviral therapy improves the ability of HDL to mediate efflux. HIV infection of macrophages decreases their ability to provide cholesterol for HDL-mediated efflux.

Current guidelines recommend a fasting lipid panel be obtained before initiating or switching therapy, and repeated at least 1 month after starting or changing an ARV regimen. In those patients with uncontrollable hyperlipidemia, switching HIV therapy should be considered, if the regimen is one linked to the observed dyslipidemia and it is refractory to conventional treatment.

Hypolipidemic therapy should be tailored to the type of dyslipidemia. For patients with triglyceride levels greater than 500 mg/dL, fibrate therapy is recommended. Gemfibrozil and fenofibrate are effective in reducing triglyceride levels in patients infected with HIV, but those who start at high triglyceride levels do not reach normal levels. Fish oil lowers triglycerides, but raises LDL, leading to no net decrease in CVD risk. Niacin is effective and has the advantage of raising HDL levels as well as lowering triglycerides, but niacin-induced insulin resistance may be problematic in HIV-infected patients. HMG-CoA reductase inhibitors (drugs of the statin class) may be better second-line agents for lowering triglycerides. As with HIV-negative patients, combination pharmacologic therapy may be required to correct extremely elevated triglyceride levels.

HMG-CoA reductase inhibitors are effective, first-line agents for the treatment of hypercholesterolemia. However, there are multiple drug-drug interactions that are important in HIV-infected patients. Some statins are metabolized by CYP3A4 which is induced or inhibited by some ARV drugs. PIs, especially ritonavir, which is used to boost levels of many HIV ARV drugs to therapeutic levels, inhibit CYP3A4 which is the major metabolic pathway for simvastatin and lovastatin. Ritonavir-based regimens

increase simvastatin levels by 5- to 32-fold, and multiple cases of rhabdomyolysis have been reported on such combinations. The combination of PI and simvastatin or lovastatin should not be used. Atorvastatin activity increases twofold, so 80 mg atorvastatin should be avoided.

Lopinavir/ritonavir increases rosuvastatin levels two- to five-fold, and tipranavir/ritonavir increases rosuvastatin levels twofold and atorvastatin levels eightfold by unknown mechanisms. If those PI combinations are used, only low doses of rosuvastatin and atorvastatin should be used. Thus, pitavastatin, pravastatin, and fluvastatin XL are recommended as first-line statins for patients on PI- or ritonavir-based HAART.

The effects of HIV protease inhibitors on glucose and lipid metabolism are shown in Table 25–5.

HIV, Antiretroviral Therapy, and Risk of Atherosclerosis

The changes in lipid and glucose metabolism seen in HIV raise the question about whether atherosclerosis increases due to HIV or its therapies. Multiple retrospective studies report an increased prevalence of CVD in HIV-infected patients. Some early studies also found an association with ARV therapy, particularly PI therapy, in addition to traditional risk factors such as age, gender, smoking, and LDL and HDL levels. A randomized trial of continuous versus intermittent therapy for HIV showed that continuous therapy has fewer cardiovascular complications. Smoking is more prevalent in those with HIV infection versus non-infected controls. Cross-sectional studies of the intima media thickness (IMT) of the carotid and femoral arteries by ultrasound have also shown that traditional CVD risk factors provide the dominant contribution to increased plaque. However, after adjusting for traditional CVD risk factors, HIV infection is an independent risk factor for increased IMT, similar in magnitude to male sex, diabetes, and smoking. As a consequence, a more aggressive to CVD prevention may be warranted.

CONCLUSION

Some of the changes in HIV infection are similar to those seen in other serious illnesses (euhormone-sick syndrome). However, many of the endocrine and metabolic changes that occur with HIV infection are unique to HIV and require careful consideration in diagnosis and treatment. In addition to the HIV disease state itself, several factors contribute to the development of these changes such as opportunistic infections, immune reconstitution, and ARV medications. As patients are living longer due to the introduction of effective ARV therapy, there is increased interest in the long-term implications of metabolic alterations such as insulin resistance, dyslipidemia, and body composition changes. Given the increased atherosclerotic disease in HIV-infected patients, more aggressive treatment of dyslipidemia is warranted, similar to the approach in diabetes or perhaps even for secondary prevention.

REFERENCES

Thyroid Function

Basilio-De-Oliveira CA. Infectious and neoplastic disorders of the thyroid in AIDS patients: an autopsy study. *Braz J Infect Dis*. 2000;4:67. [PMID: 10795071]

Chen F, Day SL, Metcalfe RA, et al. Characteristics of autoimmune thyroid disease occurring as a late complication of immune reconstitution in patients with advanced human immunodeficiency virus (HIV) disease. *Medicine*. 2005;84:98. [PMID: 15758839]

Grunfeld C, Pang M, Doerrler W, et al. Indices of thyroid function and weight loss in human immunodeficiency virus infection and the acquired immunodeficiency syndrome. *Metabolism*. 1993;42:1270. [PMID: 8412739]

Nelson M, Powles T, Zeitlin A, et al. Thyroid dysfunction and relationship to antiretroviral therapy in HIV-positive individuals in the HAART era. *J Acquir Immune Defic Syndr*. 2009;50:113. [PMID: 19092451]

Adrenal Function

Chen F, Kearney T, Robinson S, Daley-Yates PT, Waldron S, Churchill DR. Cushing's syndrome and severe adrenal suppression in patients treated with ritonavir and inhaled nasal fluticasone. *Sex Transm Infect*. 1999;75:274. [PMID: 10615321]

Honour J, Schneider M, Miller R. Low adrenal androgens in men with HIV infection and the acquired immunodeficiency syndrome. *Horm Res*. 1995;44:35. [PMID: 7649525]

Mayo J, Collazos J, Martinez E, Ibarra S. Adrenal function in the human immunodeficiency virus-infected patient. *Arch Intern Med*. 2002;162:1095. [PMID: 12020177]

Norbiato G, Bevilacqua M, Vago T, Clerici M. Glucocorticoid resistance and the immune function in the immunodeficiency syndrome. *Ann NY Acad Sci*. 1998;840:835. [PMID: 9629309]

Smolyar D, Tirado-Bernardini R, Landman R, Lesser M, Young I, Poretsky L. Comparison of 1-micro g and 250-micro g corticotropin stimulation tests for the evaluation of adrenal function in patients with acquired immunodeficiency syndrome. *Metabolism*. 2003;52:647. [PMID: 12759899]

Bone and Mineral Metabolism

Bolland MJ, Grey AB, Gamble GD, Reid IR. Clinical review: low body weight mediates the relationship between HIV infection and low bone mineral density: a meta-analysis. *J Clin Endocrinol Metab*. 2007;92:4522. [PMID: 17925333]

Brown TT, Qaqish RB. Antiretroviral therapy and the prevalence of osteopenia and osteoporosis: a meta-analytic review. *AIDS*. 2006;20:2165. [PMID: 17086056]

Earle KE, Seneviratne T, Shaker J, Shoback D. Fanconi's syndrome in HIV+ adults: report of three cases and literature review. *J Bone Miner Res*. 2004;19:714. [PMID: 15068493]

Hasse B, Ledergerber B, Egger M, et al. Antiretroviral treatment and osteonecrosis in patients of the Swiss HIV Cohort Study: a nested case-control study. *AIDS Res Hum Retroviruses*. 2004;20:909. [PMID: 15597520]

Hemkens LG, Ewald H, Santini-Oliveira M, et al. Comparative effectiveness of tenofovir in treatment-naïve HIV-infected patients: systematic review and meta-analysis. *HIV Clin Trials*. 2015;16:178. [PMID: 26395328]

O'Neill TJ, Rivera L, Struchkov V, Zaheen A, Thein HH. The effect of HIV-hepatitis C co-infection on bone mineral density and fracture: a meta-analysis. *PLoS One*. 2014;9:e101493. [PMID: 25033046]

Pinzone MR, Moreno S, Cacopardo B, Nunnari G. Is there enough evidence to use bisphosphonates in HIV-infected patients? A systematic review and meta-analysis. *AIDS Rev*. 2014;16:213. [PMID: 25300622]

Triant VA, Brown TT, Lee H, Grinspoon SK. Fracture prevalence among human immunodeficiency virus (HIV)-infected versus non-HIV-infected patients in a large U.S. healthcare system. *J Clin Endocrinol Metab*. 2008;93:3499. [PMID: 18593764]

Yombi JC, Vandercam B, Wilms D, Dubuc JE, Vincent A, Docquier PL. Osteonecrosis of the femoral head in patients with type 1 human immunodeficiency virus infection: clinical analysis and review. *Clin Rheum*. 2009;28:815. [PMID: 19277811]

Gonadal Function

Conde DM, Silva ET, Amaral WN, et al. HIV, reproductive aging, and health implications in women: a literature review. *Menopause.* 2009;16:199. [PMID: 18800017]

Cejtin HE, Kalinowski A, Bacchetti P, et al. Effects of human immunodeficiency virus on protracted amenorrhea and ovarian dysfunction. *Obstet Gynecol.* 2006;108:1423. [PMID: 17138776]

Cu-Uvin S. Effect of the menstrual cycle on virological parameters. *J Acquir Immune Defic Syndr.* 2005;38:S33. [PMID: 15867614]

Dolan SE, Grinspoon S. Androgen deficiency and the role of testosterone administration in HIV-infected women. *J Acquir Immune Defic Syndr.* 2005;38:S48. [PMID: 15867625]

Greenblatt RM, Ameli N, Grant RM, et al. Impact of the ovulatory cycle on virologic and immunologic markers in HIV-infected women. *J Acquir Immune Defic Syndr.* 1999;21:293. [PMID: 10608754]

Knapp PE, Storer TW, Herbst KL, et al. Effects of a supraphysiological dose of testosterone on physical function, muscle performance, mood, and fatigue in men with HIV-associated weight loss. *Am J Physiol Endocrinol Metab.* 2008;294:E1135. [PMID: 18430965]

Sauer MV. Sperm washing techniques address the fertility needs of HIV-seropositive men: a clinical review. *Reprod Biomed Online.* 2005;10:135. [PMID: 15705311]

Watts DH. Effect of pregnancy. *J Acquir Immune Defic Syndr.* 2005;38:S36. [PMID: 15867616]

AIDS Wasting Syndrome

Grunfeld C, Feingold KR. Metabolic disturbances and wasting in the acquired immunodeficiency syndrome. *N Engl J Med.* 1992;327:329-37. [PMID: 1620172]

Grunfeld C, Kotler DP, Dobs A, Glesby M, Bhasin S. Oxandrolone in the treatment of HIV-associated weight loss in men. *JAIDS.* 2006;41:304. [PMID: 16509313]

Kong A, Edmonds P. Testosterone therapy in HIV wasting syndrome: systematic review and meta-analysis. *Lancet Infect Dis.* 2002;2:692. [PMID: 12409050]

Schambelan M, Mulligan K, Grunfeld C, et al. Recombinant human growth hormone in patients with HIV-associated wasting. A randomized, placebo-controlled trial. Serostim Study Group. *Ann Intern Med.* 1996;125:873. [PMID: 8967667]

Storer TW, Woodhouse LJ, Sattler F, et al. A randomized, placebo-controlled trial of nandrolone decanoate in HIV-infected men with mild to moderate weight loss with recombinant human growth hormone as active reference treatment. *J Clin Endocrinol Metab.* 2005;90:4474. [PMID: 15914526]

Fat Distribution in HIV

The Study of Fat Redistribution and Metabolic Change in HIV Infection (FRAM). Fat distribution in men with HIV infection. *JAIDS.* 2005;40:121. [PMID: 16186728]

The Study of Fat Redistribution and Metabolic Change in HIV Infection (FRAM). Fat distribution in women with HIV infection. *JAIDS.* 2006;42:562. [PMID: 16837863]

Falutz J, Allas S, Blot K, et al. Metabolic effects of a growth hormone-releasing factor in patients with HIV. *N Engl J Med.* 2007;357:2359. [PMID: 18075738]

Kotler DP, Muurahainen N, Grunfeld C, et al. Effects of growth hormone on abnormal visceral adipose tissue accumulation and dyslipidemia in HIV-infected patients. *J Acquir Immune Defic Syndr.* 2004;32:239. [PMID: 15076238]

Martin A, Smith DE, Carr A, et al. Reversibility of lipoatrophy in HIV-infected patients 2 years after switching from a thymidine analogue to abacavir: the MITOX Extension Study. *AIDS.* 2004;18:1029. [PMID: 15096806]

Sivakumar T, Mechanic O, Fehmie DA, Paul B. Growth hormone axis treatments for HIV-associated lipodystrophy: a systematic review of placebo-controlled trials. *HIV Med.* 2011;12:453.

Sutinen J. Interventions for managing antiretroviral therapy-associated lipoatrophy. *Curr Opin Infect Dis.* 2005;18:25. [PMID: 15647696]

Glucose and Lipid Metabolism

Aberg JA, Gallant JE, Ghanem KG, et al. Primary care guidelines for the management of persons infected with HIV: 2013 update by the HIV Medicine Association of the Infectious Diseases Society of America. *Clin Infect Dis.* 2014;58:e1.

Currier J, Scherzer R, Bacchetti P, et al. Regional adipose tissue and lipid and lipoprotein levels in HIV-infected women. *J Acquir Immune Defic Syndr.* 2008;48:35. [PMID: 18197118]

Fichtenbaum CJ, Gerber JG, Rosenkranz SL, et al. Pharmacokinetic interactions between protease inhibitors and statins in HIV seronegative volunteers: ACTG Study A5047. *AIDS.* 2002;16:569. [PMID: 11873000]

Grunfeld C, Rimland D, Gibert CL, et al. Association of upper trunk and visceral adipose tissue volume with insulin resistance in control and HIV-infected subjects in the FRAM Study. *JAIDS.* 2007;46:283. [PMID: 18167644]

Hertel J, Struthers H, Horj CB, Hruz PW. A structural basis for the acute effects of HIV protease inhibitors on GLUT4 intrinsic activity. *J Biol Chem.* 2004;279:55147. [PMID: 15496402]

Lee GA, Mafong DD, Noor MA, et al. HIV protease inhibitors increase adiponectin levels in HIV-negative men. *J Acquir Immune Defic Syndr.* 2004;36:645. [PMID: 15097312]

Lee GA, Seneviratne T, Noor MA, et al. The metabolic effects of lopinavir/ritonavir in HIV-negative men. *AIDS.* 2004;18:641. [PMID: 15090769]

Lee GA, Rao M, Mulligan K, et al. The effects of ritonavir and amprenavir on insulin-mediated glucose disposal in healthy volunteers: two randomized, placebo-controlled, cross-over trials. *AIDS.* 2007;21:2183. [PMID: 18090045]

Mulligan K, Grunfeld C, Tai VW, et al. Hyperlipidemia and insulin resistance are induced by protease inhibitors independent of changes in body composition in patients with HIV infection. *J Acq Immune Defic Syndr Hum Retrovirol.* 2000;23:35. [PMID: 10708054]

Noor MA, Lo J, Mulligan K, et al. Metabolic effects of indinavir in healthy HIV-seronegative men. *AIDS.* 2001;15:F11. [PMID: 11399973]

Noor MA, Parker RA, O'Mara E, et al. The effects of HIV protease inhibitors atazanavir and lopinavir/ritonavir on insulin sensitivity in HIV-seronegative healthy adults. *AIDS.* 2004;18:2137. [PMID: 15577646]

Pao VY, Lee GA, Taylor S, et al. The protease inhibitor combination lopinavir/ritonavir does not decrease insulin secretion in healthy, HIV-seronegative volunteers. *AIDS.* 2010;24:265. [PMID: 19890203]

Phillips AN, Carr A, Neuhaus J, et al. Interruption of antiretroviral therapy and risk of cardiovascular disease in persons with HIV-1 infection: exploratory analyses from the SMART trial. *Antivir Ther.* 2008;13:177. [PMID: 18505169]

Siegel MO, Borkowska AG, Dubrovsky L, et al. HIV infection induces structural and functional changes in high density lipoproteins. *Atherosclerosis.* 2015;243:19. [PMID: 26343868]

van der Valk M, Kastelein JJ, Murphy RL, et al. Nevirapine-containing antiretroviral therapy in HIV-1 infected patients results in an antiatherogenic lipid profile. *AIDS.* 2001;15:2407. [PMID: 11740191]

Wohl D, Scherzer R, Heymsfield S, et al. The associations of regional adipose tissue with lipid and lipoprotein levels in HIV-infected men. *JAIDS.* 2008;48:44. [PMID: 18360291]

HIV, Antiretroviral Therapy, and the Risk of Atherosclerosis

Currier JS, Kendall MA, Zackin R, et al. Carotid artery intima-media thickness and HIV infection: traditional risk factors overshadow impact of protease inhibitor exposure. *AIDS.* 2005;19:927. [PMID: 15905673]

DAD Study Group, Friis-Moller N, Reiss P, et al. Class of antiretroviral drugs and the risk of myocardial infarction. *N Engl J Med.* 2007;356:1723. [PMID: 17460226]

Grinspoon SK, Grunfeld C, Kotler DP, et al. Initiative to decrease cardiovascular risk and increase quality of care for patients living with HIV/AIDS: executive summary. *Circulation.* 2008;118:198. [PMID: 19455012]

Grunfeld C, Delaney JA, Wanke C, et al. Preclinical atherosclerosis due to HIV infection: carotid intima-medial thickness measurements from the FRAM study. *AIDS.* 2009;23:1841. [PMID: 19455012]

Hsue PY, Lo JC, Franklin A, et al. Progression of atherosclerosis as assessed by carotid intima-media thickness in patients with HIV infection. *Circulation.* 2004;109:1603. [PMID: 15023877]

Geeta Lal, MD, MSc, FRCS(C), FACS and Orlo H. Clark, MD

ACTH	Adrenocorticotropic hormone		**PTH**	Parathyroid hormone
CASR	Calcium-sensing receptor		**RAI**	Radioactive iodine
CT	Computed tomography		**RAIU**	Radioactive iodine uptake
FHH	Familial hypocalciuric hypercalcemia		**RFA**	Radiofrequency ablation
FNAB	Fine-needle aspiration biopsy		**TACE**	Transarterial chemoembolization
FU	Fluorouracil		**SPECT**	Single photon emission computed tomography
MEN	Multiple endocrine neoplasia		**TSH**	Thyroid stimulating hormone
MRI	Magnetic resonance imaging		**VIP**	Vasoactive intestinal peptide
MTC	Medullary thyroid cancer		**ZES**	Zollinger-Ellison syndrome
NIH	National Institutes of Health			
PHPT	Primary hyperparathyroidism			

INTRODUCTION

Many endocrine diseases are appropriately managed surgically. The details of clinical presentation, diagnosis, and medical management are discussed in other sections of this book. This chapter provides an overview of the principles involved in the surgical therapy for these conditions. The indications for surgical intervention, the relevant procedures, and their risks and benefits are discussed.

THE THYROID GLAND

EMBRYOLOGY AND ANATOMY

The thyroid gland arises in the midline as an endoderm-derived pharyngeal diverticulum around the third week of gestation. The paired median thyroid anlages then descend from their origin at the base of the tongue (foramen cecum) and ultimately form a bilobed thyroid gland anterolateral to the trachea and larynx. The thyroid lobes are connected just below the cricoid cartilage by an isthmus. The connection to the foramen cecum—the thyroglossal duct—separates and is partially resorbed by the sixth week of gestation. Its distal remnant forms the pyramidal lobe. The calcitonin-producing C cells are neuroectodermal in origin, arise from

the fourth branchial pouch, and are located in the lateral superoposterior aspect of the gland.

A number of embryologic or developmental abnormalities of the thyroid have been described and are related to the absence or mutations of thyroid differentiation factors, including thyroid transcription factors 1 and 2 (TTF-1 and TTF-2) and transcription factor Pax 8. Thyroglossal duct cysts are usually found in the midline, just inferior to the hyoid bone. A lingual thyroid results from maldescent of the median thyroid anlage and is often accompanied by agenesis of other thyroid tissue. Rests of thyroid tissue may be found anywhere in the central compartment of the neck, including the anterior mediastinum. *Tongues* of thyroid tissue are often seen to extend off the lower thyroid poles, particularly in large goiters. In contrast to the above, thyroid tissue in the lateral neck lymph nodes (lateral aberrant thyroid) almost always represents metastatic thyroid cancer and is not a developmental abnormality.

The adult thyroid gland is reddish-brown in color and weighs approximately 20 g. The thyroid gland is supplied by paired superior and inferior thyroid arteries. The former arise from the external carotid artery and the latter from the thyrocervical trunk. A thyroid ima artery arises directly from the aorta or innominate artery in approximately 2% of individuals and enters the isthmus, replacing an absent inferior artery.

TABLE 26–1 **Definitions of various thyroid resections.**

Procedure	Description
Nodulectomy or lumpectomy	Removal of lesion with minimal surrounding tissue
Partial thyroidectomy	Removal of lesion and larger rim of normal tissue
Subtotal thyroidectomy	Bilateral removal of >50% of each lobe and an isthmusectomy
Lobectomy or hemithyroidectomy	Complete removal of a lobe and isthmus
Near-total thyroidectomy	Complete removal of one lobe and isthmus and all but 1 g (1 cm) of the contralateral lobe (tissue near ligament of Berry)
Total thyroidectomy	Complete removal of both thyroid lobes, isthmus, and pyramidal lobe

The thyroid is drained by three sets of veins: the superior, middle, and inferior thyroid veins. The first two drain into the internal jugular vein, whereas the last drains into the innominate veins. Both recurrent laryngeal nerves arise from their respective vagus nerves and enter the larynx at the level of the cricothyroid articulation, posterior to the cricothyroid muscle. The left recurrent laryngeal nerve recurs around the ligamentum arteriosum and ascends to the larynx in the tracheoesophageal groove. The right recurrent laryngeal nerve recurs around the subclavian artery and runs 1 to 2 cm lateral to the tracheoesophageal groove at the level of the clavicle and courses obliquely to the larynx. The superior laryngeal nerves also arise from corresponding vagus nerves and divide into internal and external branches. The former provides sensation to the larynx and the latter innervates the cricothyroid muscles. A description of parathyroid embryology and anatomy is presented in the next section.

Prior to discussing indications for thyroidectomy, it is prudent to clarify the definitions of various thyroid resections. A description of these terms is presented in Table 26–1. Of note, nodulectomies are rarely performed, and most surgeons agree that thyroid lobectomy constitutes a minimum resection for any nodule that could be malignant.

INDICATIONS FOR SURGERY

DEVELOPMENTAL THYROID ABNORMALITIES

Thyroglossal duct remnants may become symptomatic, forming cysts, abscesses, and fistulae. There is also a 1% risk of thyroid cancer development in thyroglossal duct cysts. Most are papillary carcinomas, but very rarely a squamous cell carcinoma may develop. Medullary thyroid cancers do not occur at this site.

Treatment of thyroglossal duct remnants is surgical and consists of the Sistrunk procedure, which involves removal of the cyst and duct up to the foramen cecum. Because the duct may pass anterior to, posterior to, or through the hyoid bone, it is generally recommended that the midsection of this bone is also resected in order to reduce the risk of recurrence. Surgery may also be needed for enlarged lingual thyroid tissue causing symptoms such as choking, dysphagia, airway obstruction, and hemorrhage. Prior to resection, care must be taken to determine whether the patient has any other functioning thyroid tissue, usually via a thyroid scan or ultrasound.

HYPERTHYROIDISM

Hyperthyroidism may result from increased thyroid hormone production or release of stored thyroid hormone following injury to the thyroid gland such as that seen in subacute thyroiditis. This is an important distinction, because the former group of disorders leads to increased radioactive iodine uptake (RAIU), whereas the latter are associated with low or normal RAIU. The most common causes of hyperthyroidism are diffuse toxic goiter (Graves' disease), toxic multinodular goiter or a single toxic nodule, all of which cause increased RAIU. Rarer causes of hyperthyroidism with increased RAIU include a TSH–secreting tumor and a hydatidiform mole. Causes of hyperthyroidism without increased RAIU include subacute thyroiditis, excessive ingestion of medicinal thyroid hormone, struma ovarii, and thyroid hormone–secreting metastatic thyroid cancer. These latter conditions present with the usual symptoms and signs of hyperthyroidism but lack the extrathyroidal manifestations of Graves' disease such as ophthalmopathy, pretibial myxedema, and thyroid acropachy.

Diagnostic Tests

Hyperthyroidism is characterized by a suppressed TSH in the presence of elevated free T_4 levels. If free T_4 is normal, free T_3 levels should be measured because they are often elevated early in the course of disease. RAIU can be used to distinguish the various causes of hyperthyroidism. Approximately 90% of patients with Graves' disease have elevated levels of thyroid-stimulating antibodies or immunoglobulins (thyroid-stimulating immunoglobulin).

Management of Hyperthyroidism

Hyperthyroidism may be treated medically with antithyroid medications, but relapse is common when the medical treatment is discontinued. Ablation of the thyroid gland with RAI is the mainstay of definitive treatment in North America for patients over 30 years of age. A dose of 8 to 12 mCi of I-131 taken orally is generally required. However, it is associated with a prolonged latency period before effective action, a slightly increased risk of future benign and malignant thyroid tumors, hyperparathyroidism, worsening ophthalmopathy (particularly in smokers), and unavoidable hypothyroidism (3% per year after the first year, independently of dosage). Furthermore, only about 50% of patients treated with RAI are euthyroid at 6 months. It is contraindicated in pregnant women, of concern in children, and should be avoided in women wishing to become pregnant for up to 1 year after treatment. Surgery overcomes many of the problems

associated with RAI and achieves rapid control of hyperthyroidism with minimal side effects in experienced hands.

Absolute indications for thyroidectomy in patients with Graves' disease include biopsy-proven suspicious or cancerous nodules, local compressive symptoms, severe reactions to antithyroid medications, reluctance to undergo RAI ablation, or fear of recurrence after RAI. Women who want to become pregnant after treatment or who develop side effects from antithyroid drugs during pregnancy are also candidates for thyroidectomy, as are children. Relative indications for thyroidectomy include patients in whom rapid control of the disease is desired; poorly compliant patients; and patients with severe ophthalmopathy, very large goiters, or low RAIU.

Preoperative Preparation

Patients are usually treated with antithyroid medications to render them euthyroid and to prevent the risk of thyroid storm. Propylthiouracil (100-200 mg three times daily) or methimazole (10-20 mg twice daily and then once daily) are the most commonly used medications and are continued up to the day of surgery. In patients who develop agranulocytosis as a complication of these medications, surgery should be deferred until granulocyte counts reach 1000 cells/uL. In addition, patients are also often treated with propranolol (10-40 mg four times daily) to control the adrenergic effects of hyperthyroidism. Relatively large doses may be necessary because of increased catabolism of the drug. Lugol's solution (iodine and potassium iodide) or saturated solution of potassium iodide is also generally started about 7 to 10 days preoperatively to reduce the vascularity of the gland and decrease the risk of precipitating thyroid storm.

Extent of Surgery

The extent of surgery done has traditionally depended on multiple factors. For those patients who experienced severe complications with antithyroid drugs, those who wanted to eliminate the risk of recurrence, and those with coexistent carcinoma or severe ophthalmopathy, total or near-total thyroidectomy was recommended. However, total thyroidectomy results in inevitable hypothyroidism and is potentially associated with higher complication rates when compared to lesser procedures. Thus, the procedure typically preferred by most surgeons for treating patients with Graves' disease was a subtotal thyroidectomy. This was accomplished by bilateral subtotal excision or by unilateral lobectomy and isthmusectomy with subtotal contralateral resection (Hartley-Dunhill procedure). Provided an adequate remnant was left behind, the procedure provided rapid relief of thyrotoxicosis, while maintaining a euthyroid state, without the need for thyroid hormone. However, it was difficult to define what constituted an adequate remnant, although most surgeons considered a 4- to 7-g remnant sufficient in adults. Rates of postoperative hypothyroidism are variable and are primarily determined by remnant size, thyroid antibody titer, and whether hypothyroidism was reported to be subclinical or overt. Remnants less than 4 g were associated with a more than 50% risk of hypothyroidism, and those more than 8 g were associated with recurrence rates of 15%. Therefore, many surgeons now consider avoidance of recurrence, rather than attainment of euthyroidism, as the goal for

thyroidectomy in Graves' disease. Based on the current evidence, recently published guidelines from the American Thyroid Association (ATA) and the American Association of Clinical Endocrinologists (AACE) recommend total or near-total thyroidectomy as the procedure of choice for the surgical management of Graves' disease. Patients who experience recurrence after surgery are usually treated with RAI. Patients with hyperthyroidism secondary to toxic multinodular goiter are managed similar to those with Graves' disease. Subtotal thyroidectomy is generally advised, however, if no normal remnant can be left, a near-total or total thyroidectomy is recommended. Patients with a solitary toxic nodule larger than 3 cm in size are best treated by ipsilateral lobectomy. Patients with smaller *hot* nodules may be managed with antithyroid medication and RAI ablation. Percutaneous ablation of the nodule using ethanol injection has also been proposed, however, long-term outcomes are unknown.

THYROIDITIS

Thyroiditis, an inflammatory disorder of the thyroid gland, may be classified as acute, subacute, or chronic. Acute suppurative thyroiditis is diagnosed by fine-needle aspiration for cytology, smear, Gram stain, and culture and treated by incision and drainage and antibiotics. Acute thyroiditis is usually self-limited. Thyroidectomy, however, is occasionally needed for clinically coexistent suspicious nodules or cancer, local compressive symptoms, or persistent infection. Recurrent acute thyroiditis is often due to a persistent pyriform sinus fistula. Resection of the entire tract is necessary to prevent recurrence. Subacute and chronic thyroiditis is usually managed medically, but surgical resection is occasionally needed for local symptoms or relapse.

GOITER (NONTOXIC)

Goiters may be diffuse, uninodular, or multinodular. Thyroidectomy is indicated for enlarging goiters despite thyroxine suppression, those causing compressive symptoms (choking, dysphagia, hoarseness, dyspnea, or a positive Pemberton's sign—dilation of neck veins and facial erythema on elevation of arms), nodules deemed suspicious by ultrasound, and goiters containing biopsy-proven suspicious or cancerous nodules. Goiters with a significant substernal component (>50%) are considered a relative indication for thyroidectomy. A total or near-total thyroidectomy is the treatment of choice.

THYROID NODULES

Approximately 4% of the North American population develops thyroid nodules. However, the incidence of clinical thyroid cancer is much lower (about 40 patients per million), thus making it imperative to determine which nodules require treatment with thyroidectomy. A thyroid nodule is more likely to be malignant if the patient has a history of therapeutic radiation to the head and neck (6.5-3000 cGy to the thyroid), a family history of thyroid cancer, Cowden syndrome, or MEN 2, and a personal history of thyroid cancer. Other features suggesting cancer include male sex,

young (<20) or old (>70) age; a solitary, *cold* solid (or mixed solid-cystic), hard nodule, a nodule with microcalcifications or irregular edges on ultrasonography; the presence of ipsilateral palpable nodes or vocal cord palsy; and a fine-needle aspiration biopsy (FNAB) suspicious for cancer. Approximately 40% of individuals with a history of therapeutic radiation exposure or a family history of thyroid cancer have a thyroid cancer. The cancer is in the index nodule in 60% and in other nodules in 40% of patients.

Diagnostic Tests

Patients with thyroid nodules should be evaluated with TSH measurement and FNAB. The Besthesda Classification of Thyroid FNAB and the associated risk of malignancy for each category are discussed elsewhere in this book. If the biopsy suggests a follicular neoplasm, and the TSH level is suppressed or is in the low-normal range, an RAI scan should be performed to rule out a *hot* nodule as the risk of malignancy is very low in this setting.

Management

Nodules with any of the worrisome features mentioned above should be removed with—at a minimum—an ipsilateral lobectomy and isthmusectomy. If FNAB confirms a malignancy, total thyroidectomy may be needed based on the final pathology. Patients with benign nodules on FNAB may be treated with exogenous thyroxine in areas with a high prevalence of iodine deficiency, although this practice is not recommended in iodine sufficient populations. Nodules that continue to enlarge or become symptomatic during follow-up may be re-biopsied or removed via a thyroidectomy.

THYROID CANCER

Malignant thyroid tumors include differentiated lesions (which arise from follicular cells); MTC; undifferentiated or anaplastic cancers; and other rare tumors such as lymphomas, squamous cell carcinomas, sarcomas, teratomas, plasmacytomas, paragangliomas, and metastatic thyroid cancers (from melanoma or from breast, kidney, lung, and other head and neck tumors).

1. DIFFERENTIATED THYROID CANCER

This group includes papillary, follicular variant of papillary, follicular, and Hürthle cell tumors. Hürthle cell carcinoma has been considered a subtype of follicular carcinoma by some investigators and a unique differentiated thyroid cancer of follicular cell origin by others.

The characteristics of these tumors are depicted in Table 26–2. The biologic behavior of the follicular variant of papillary thyroid cancer is similar to that of papillary carcinoma. In follicular and Hürthle cell tumors, the diagnosis of malignancy can only be made by the presence of capsular, blood vessel, or lymphatic invasion or when lymph node or distant metastases are present. Familial non-MTCs, especially if there is a family history of more than two affected relatives, have been shown to be more aggressive than their sporadic counterparts.

TABLE 26–2 Characteristics of differentiated thyroid cancers.

Feature	Papillary	Follicular	Hürthle
Frequency	80%	10%-20%	3%-5%
Age group (y)	20-30	40-50	50-60
Multicentric	85%	10%	30%
Lymph node metastases	30%-40%	10%	25%
Distant metastases	2%-14%	33%	15%
RAI uptake	70%	80%	10%
Prognosis (10-y survival)	95%	85%	65%

Surgical Treatment

Papillary carcinomas can be easily diagnosed by FNAB (Bethesda category 6). Patients with occult or minimal papillary carcinomas (<1 cm) have an excellent prognosis and are adequately treated by lobectomy. Lobectomy alone is also considered sufficient treatment for low-risk, unifocal, intrathyroidal cancers in the absence of a history of head and neck irradiation or a family history of thyroid cancer or cervical node metastases. Patients with high-risk cancers (determined by AGES, AMES, or TNM classification; Tables 26–3, 26–4, and 26–5), bilateral cancers, clinically apparent nodal or distant metastases are best treated by total or near-total thyroidectomy. Considerable debate has existed regarding optimal treatment for low-risk differentiated thyroid cancer larger than 1 cm and less than 4 cm without extrathyroidal extension or lymph node metastases.

Proponents of total thyroidectomy argued that this procedure is advantageous for several reasons: (1) RAI can be used to diagnose and treat recurrent or metastatic disease, (2) serum thyroglobulin becomes a sensitive indicator of recurrent disease, (3) the procedure eliminates the risk of growth of occult cancer in the contralateral lobe and reduces the risk of recurrence, (4) it decreases the 1% risk of progression to undifferentiated cancer, and (5) it decreases the risk of reoperation in case of central neck recurrence. On the other hand, those favoring thyroid lobectomy

TABLE 26–3 AGES system of classifying high-risk patients.

Variable	Description
Age	Women older than 50 y Men older than 40 y
Grade	Poorly differentiated Fibrous stroma Insular, mucoid, and tall cell variants
Extent	Invasive into adjacent tissues or distant metastases
Size	Tumor with a maximum diameter of >4 cm

Reproduced with permission from Clark OH, Duh Q-Y. *Textbook of Endocrine Surgery*. Philadelphia: Saunders; 1997.

TABLE 26–4 AMES system for classifying high-risk patients.

Variable	Description
Age	Men >40 y, women >50 y
Metastases	Distant metastases
Extent	Invasion of adjacent tissues
Size	>5 cm

noted that (1) total thyroidectomy is associated with a higher complication rate and that 50% of local recurrences can be cured with surgery, (2) less than 5% of recurrences occur in the thyroid bed, (3) multicentricity within the thyroid is not clinically significant, and (4) the prognosis is excellent in patients with low-risk tumors undergoing lobectomy.

Retrospective data indicate that the recurrence rate for patients with low-risk differentiated thyroid cancer is 10% and that the overall mortality rate is approximately 4% at 10 to 20 years. However, among patients who have recurrences, 33% to 50% die from thyroid cancer. Recent studies involving more than 50,000 patients with papillary cancers demonstrated that total thyroidectomy is associated with improved recurrence and survival rates even in patients with low-risk tumors. Furthermore, the most important information regarding risk for recurrence is available only postoperatively. Therefore, 2009 guidelines from the American Thyroid Association also recommend near-total or total thyroidectomy for thyroid cancers larger than 1 cm unless there

TABLE 26-5 Tumor (T), lymph node (N), and distant metastasis (M) classification and staging of thyroid cancer.

Definition of Primary Tumor (T)

Papillary, Follicular, and Hurthle Cell Thyroid Carcinoma

T Category	T Criteria
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Tumor ≤2 cm in greatest dimension limited to the thyroid
T1a	Tumor ≤1 cm in greatest dimension limited to the thyroid
T1b	Tumor >1 cm but ≤2 cm in greatest dimension limited to the thyroid
T2	Tumor >2 cm but ≤4 cm in greatest dimension limited to the thyroid
T3	Tumor >4 cm limited to the thyroid, or gross extrathyroidal extension invading only strap muscles
T3a	Tumor >4 cm limited to the thyroid
T3b	Gross extrathyroidal extension invading only strap muscles (sternohyoid, sternothyroid, thyrohyoid, or omohyoid muscles) from a tumor of any size
T4	Includes gross extrathyroidal extension
T4a	Gross extrathyroidal extension invading subcutaneous soft tissues, larynx, trachea, esophagus, or recurrent laryngeal nerve from a tumor of any size
T4b	Gross extrathyroidal extension invading prevertebral fascia or encasing the carotid artery or mediastinal vessels from a tumor of any size

Note: All categories may be subdivided: (s) solitary tumor and (m) multifocal tumor (the largest tumor determines the classification).

Definition of Distant Metastasis (M)

M Category	M Criteria
M0	No distant metastasis
M1	Distant metastasis

Definition of Regional Lymph Node (N)

N Category	N Criteria
NX	Regional lymph nodes cannot be assessed
N0	No evidence of locoregional lymph node metastasis
N0a	One or more cytologically or histologically confirmed benign lymph nodes
N0b	No radiologic or clinical evidence of locoregional lymph node metastasis
N1	Metastasis to regional nodes
N1a	Metastasis to level VI or VII (pretracheal, paratracheal, or prelaryngeal/Delphian, or upper mediastinal) lymph nodes. This can be unilateral or bilateral disease
N1b	Metastasis to unilateral, bilateral, or contralateral lateral neck lymph nodes (levels I, II, III, IV, or V) or retropharyngeal lymph nodes

AJCC Prognostic Stage Groups

Differentiated

When age at Diagnosis is...	And T is...	And N is...	And M is...	Then the Stage Group is...
<55 years	Any T	Any N	M0	I
<55 years	Any T	Any N	M1	II
≥55 years	T1	N0/NX	M0	I
≥55 years	T1	N1	M0	II
≥55 years	T2	N0/NX	M0	I
≥55 years	T2	N1	M0	II
≥55 years	T3a/T3b	Any N	M0	II
≥55 years	T4a	Any N	M0	III
≥55 years	T4b	Any N	M0	IVA
≥55 years	Any T	Any N	M1	IVB

Reproduced with permission from Amin MB, Edge SB, Greene FL, et al. *AJCC Cancer Staging Manual*, 8th ed. New York: Springer; 2017.

were contraindications to this procedure. However, additional studies since then (including a multivariate analysis) have shown that there is no difference in survival based on initial surgical procedure when adjustments were made for complexity/risk and comorbid conditions. In addition, the declining use of RAI ablation in general and greater use of ultrasound and Tg measurements for detecting recurrences led to a revision of the recommendations. The 2015 ATA guidelines indicate that thyroid lobectomy alone may be adequate as initial therapy for low-risk papillary or follicular cancers. Criteria for recommending bilateral surgery include radiation exposure, family history of DTC, older age or the presence of contralateral nodules.

It is not usually possible to distinguish follicular and Hürthle cell carcinomas from corresponding adenomas preoperatively. For patients with lesions that are consistent with follicular or Hurhle cell neoplasm or suspicious for the same (Bethesda category 4) a lobectomy is performed if there are no obvious signs of cancer at surgery (lymphadenopathy or extrathyroidal invasion) as more than 80% of these tumors are benign. If final pathology confirms cancer, a completion thyroidectomy is usually recommended, except in those patients with minimally invasive tumors. Total or near-total thyroidectomy may be performed at the initial operation in individuals with large (>4 cm) tumors, particularly if nodules are present bilaterally. Of note, molecular tests can now be performed on FNAB material to determine the risk of malignancy and guide recommendations for surgical therapy. For nodules classified as "suspicious for cancer" (Bethesda catergory 5), ipsilateral lobectomy with frozen section is performed. Completion thyroidectomy may be indicated based on the diagnosis and extent of the primary lesion.

Postoperative Treatment

After surgical therapy, patients are risk stratified based on current ATA guidelines. Low-risk patients include those with intrathyroidal papillary cancers without vascular invasion, small volume nodal metastases (N_0) or less than 5 N1 micrometastases, (less than 2 mm in largest dimension), encapsulated, intrathyroidal follicular variant of papillary cancers, intrathyroidal well-differentiated follicular cancer with capsular or minor vascular (<4 vessels) invasion and BRAF-mutated or wild-type papillary microcarcinomas. The intermediate-risk category includes patients with microscopic invasion of tumor onto perithyroid soft tissues, vascular invasion, uptake outside the thyroid bed at remnant ablation, aggressive histologies, clinical N2 or less than 5 pathologic N2 with all nodes less than 3 cm, intrathyroidal papillary cancer with a primary 1 to 4 cm that is BRAF-mutated (if known) and multifocal papillary microcarcinoma with extrathyroidal extention and BRAF-mutated (if known). The high-risk category includes patients with macroscopic extrathyroidal extension, incomplete tumor resection, distant metastases, large volume lymph node involvement, follicular cancer with extensive vascular invasion (>4 foci or extracapsular vascular invasion). Long-term cohort studies by Mazzaferri and Jhiang and others demonstrate that postoperative RAI therapy reduces recurrence rates and leads to improved survival even in patients with low-risk tumors. Despite that, in general, there has been a trend toward reduced RAI ablation rates due to concerns regarding risks of secondary malignancies and the general good prognosis of DTC. According to recently revised guidelines from the American Thyroid Association, RAI remnant ablation is not routinely recommended for low-risk patients or those with unifocal or multifocal papillary microcarcinoma in the absence of other adverse features that may modulate risk. RAI adjuvant therapy should be considered after total thyroidectomy in intermediate-risk patients and is routinely recommended for high-risk patients. Thyroxine therapy is not only important as replacement but is given in higher doses to suppress TSH levels to less than 0.1 μU/L in high-risk patients and 0.1 to 0.5 mU/L in intermediate risk patients, thus decreasing the growth stimulus for thyroid cancer cells. For low risk patients with undetectable Tg levels (with or without remnant ablation), or those having had a lobectomy alone, the TSH can be maintained in the low normal reference range (0.5-2 mU/L). If the Tg level is detectable, a TSH target of 0.1 to 0.5 mU/L is recommended. The need for TSH suppression must be balanced with the risks of subclinical thyrotoxicosis, such as exacerbation of angina in patients with ischemic heart disease, increased risk for atrial fibrillation, and osteoporosis in postmenopausal women.

2. MEDULLARY THYROID CANCER

This tumor comprises 7% of thyroid malignancies but accounts for approximately 17% of thyroid cancer–related deaths. It arises from the parafollicular (C cells) of the thyroid, which are derived from the neural crest and secrete calcitonin. MTC may be sporadic (75%) or occur in the setting of MEN 2A, MEN 2B, and familial non-MEN MTC. In the hereditary setting, the tumors are often bilateral and multicentric (90%). Approximately 50% of patients with MTC have nodal metastases in the central or lateral neck at presentation. All patients presenting with MTC should be screened for pheochromocytomas, hyperparathyroidism, and mutations of the *RET* proto-oncogene.

The ATA published revised guidelines for the management of medullary cancers in 2015. In terms of surgical therapy, pheochromocytomas should be treated prior to thyroidectomy to avoid precipitating an intraoperative hypertensive crisis. Total thyroidectomy and bilateral central compartment lymphadenectomy is the treatment of choice due to the high incidence of multicentric disease and ineffectiveness of I-131 therapy for this tumor.

3. UNDIFFERENTIATED (ANAPLASTIC) THYROID CANCER

This tumor type constitutes approximately 1% of all thyroid cancers and is the most aggressive variant. The peak incidence is in the seventh decade of life. Lymph node involvement is early and common (84%), as is local invasion into the larynx, vocal cords, recurrent laryngeal nerve, esophagus, and major vessels. Distant metastases are present in about 75% of patients. Per guidelines from the ATA, in patients without advanced disease, total or near-total thyroidectomy can be performed for cure in a minority of cases. If extrathyroidal disease is present and en bloc resection may be considered if all gross disease can be removed. External beam

radiation and chemotherapy are usually recommended as adjuvant treatment. They can also be used in the neoadjuvant setting in patients with unresectable disease. A tracheostomy is indicated for impending airway loss.

4. THYROID LYMPHOMA AND METASTASES

Most thyroid lymphomas are of the non-Hodgkin B cell type and often develop in patients with longstanding chronic lymphocytic thyroiditis. Chemotherapy and radiotherapy form the mainstay of treatment, with thyroidectomy and node dissection reserved for palliation of airway obstruction or in patients who do not respond to the above therapies. The thyroid gland is a rare site of metastases from tumors of the kidney, lung, breast, and melanoma. In selected patients, thyroidectomy may increase survival.

5. MANAGEMENT OF LYMPH NODES IN THYROID CANCER

Clinically inapparent lymph node metastases may be present in 20% to 90% of papillary thyroid cancers, and a smaller proportion of patients with other tumor subtypes. Several retrospective studies have suggested that lymph node metastases do not have a significant effect on survival in papillary thyroid cancers. Others report that lymph node metastases are associated with poorer survival but only in patients with follicular cancers and papillary cancer greater than 45 years of age. In patients with matted nodes or extranodal invasion, the recurrence rates are higher and the prognosis is worse.

The lymph node regions of the neck are depicted in Figure 26–1. A preoperative neck ultrasound is recommended for patients with malignant findings on FNAB to evaluate the contralateral lobe and lymph node regions. All clinically apparent disease in the central neck should be removed at the time of total thyroidectomy to improve clearance of disease from the central neck. Bilateral prophylactic central node dissection (level VI) may also improve survival compared to historic controls and decrease the risk of central neck recurrence. Other studies, however, show that central neck dissections are associated with a higher risk of hypoparathyroidism and recurrent nerve injury rates without any reduction in recurrence rates. Given the above, recently published ATA guidelines do not recommend routine prophylactic lymph node dissections but indicate that they should be considered in patients with advanced primary tumors (T_3 or T_4), clinically involved lateral neck nodes (N1b) or if the information may be useful in planning additional therapy. Prophylactic central neck clearance should also be considered for Hürthle cell cancers which tend to be less RAI-responsive, provided they can be performed by experienced surgeons without increased rates of hypoparathyroidism and recurrent nerve injury.

Papillary and Hürthle cell cancers can also metastasize to lymph nodes in the lateral and posterior compartments (II, III, IV, and V). In patients with biopsy-proven nodal involvement, either clinically or on preoperative ultrasound examination, an ipsilateral modified radical neck dissection is recommended, particularly if these
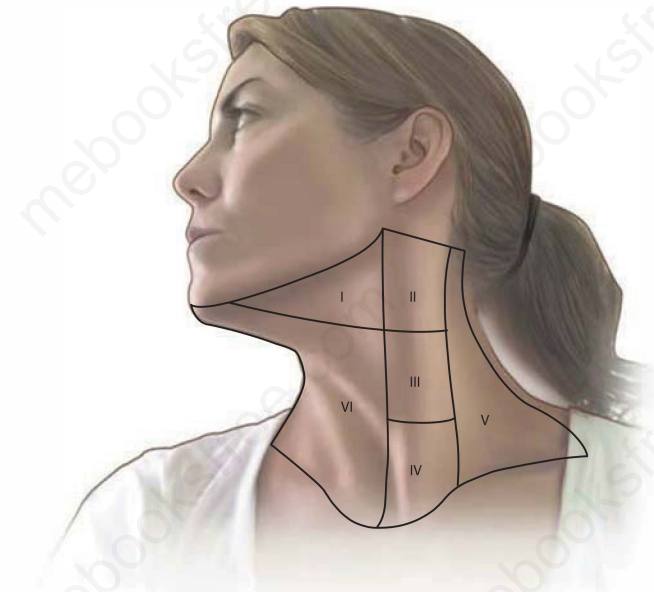
FIGURE 26–1   Lymph node regions of the neck. Level I, submandibular; levels II, III, and IV, upper, middle, and lower jugular nodes; level V, posterior triangle nodes; and level VI, central compartment nodes.

patients may fail RAI therapy based on lymph node number, size, or aggressive primary tumor histology. This procedure removes all the fibrofatty and lymph node tissue while preserving the internal jugular vein, the spinal accessory nerve, the cervical sensory nerves, and the sternocleidomastoid muscle and may reduce the risk of recurrence and subsequent mortality. Because thyroid cancers rarely metastasize to compartment I, these nodes are not routinely removed during a modified radical neck dissection.

In case of MTC, prophylactic central neck lymph node clearance is recommended because patients with these tumors have a worse prognosis and do not take up RAI. In patients with involved lateral neck nodes or calcitonin levels more than 400 pg/mL, additional imaging to include a neck and chest CT and a triple phase liver CT or contrast-enhanced MRI should be obtained to assess for metastatic disease. In patients with no distant disease, an ipsilateral or bilateral lateral neck dissection (levels IIA, III, IV, and V) is indicated. Less aggressive neck surgery should be considered to preserve speech and swallowing while maintaining locoregional control in patients with limited metastatic disease. The role of prophylactic lateral neck dissection is controversial. Some groups favor this procedure if central neck lymph nodes are involved or if the primary tumor is larger than 1.5 cm.

6. RECURRENT AND METASTATIC THYROID CANCER

Thyroglobulin levels are generally undetectable in patients following total thyroidectomy and RAI ablation. Patients with suppressed thyroglobulin levels less than 0.2 ng/mL or TSH-stimulated

levels less than 1 ng/mL in the absence of thyroglobulin antibodies are deemed to have had an excellent response to treatment. Patients with suspected recurrence should be imaged with neck ultrasound, CT, or MRI. In general, surgical extirpation is recommended for patients with macronodular disease (>1 cm in diameter), followed by RAI therapy and TSH suppression. Percutaneous ethanol ablation or radiofrequency ablation (RFA) of recurrent nodal disease may also be used in selected cases. Thyroid cancers may metastasize to the lungs, bone, liver, and brain. Microscopic lung metastases in patients with papillary and follicular thyroid cancer are best treated with RAI and macroscopic disease surgically although RAI may be useful if the latter are iodine avid. External beam radiation is useful for unresectable, locally invasive or recurrent disease and to treat bony metastases with minimal to no RAI uptake. Therapy with kinase inhibitors should be considered in RAI-refractory DTC patients with metastatic, rapidly progressive, symptomatic and/or imminently threatening disease which cannot be locally controlled by other means.

CONDUCT OF THYROIDECTOMY

A preoperative laryngeal examination is recommended in all patients who report voice changes, have a history of prior neck or chest surgery and those known to have extensive disease at presentation. A 4- to 5-cm incision is placed in or parallel to a natural skin crease 1 cm below the cricoid cartilage (Figure 26–2A). The subcutaneous tissue and platysma are divided, subplatysmal flaps are raised (Figure 26–2B), and the strap muscles are separated in the midline from the thyroid cartilage to the suprasternal notch. Initial dissection is begun in the midline by identification of delphian nodes and the pyramidal lobe, followed by division of the fascia just cephalad to the isthmus. The trachea is then cleared just caudal to the isthmus. The thyrothymic ligaments and the inferior thyroid veins are ligated and divided. The side with the dominant or suspicious mass is approached first. In case of a proposed lobectomy or the absence of cancer, the isthmus is divided. Tissues are swept lateral to the thyroid by blunt dissection, and the middle thyroid veins are ligated and divided (Figure 26–2C). The superior pole vessels are then individually ligated and divided low on the thyroid gland to decrease the risk of injury to the external branch of the superior laryngeal nerve (Figure 26–2D). The recurrent laryngeal nerve and the superior parathyroid gland are identified at the level of the cricoid cartilage. Once this is accomplished, the ligament of Berry is divided (Figure 26–2E), and the thyroid is sharply dissected off the trachea. Of note, the current ATA guidelines advise routine identification of the recurrent laryngeal nerve in all cases, although routine use of neuromonitoring is not recommended. The same procedure is repeated on the other side for a total thyroidectomy.

Several approaches to minimally invasive thyroidectomy such as mini-incision procedures, video-assisted thyroidectomy and total endoscopic thyroidectomy via supraclavicular, axillary incisions, chest, and breast approaches have been described (Figure 26–3). The endoscopic approaches can also be performed with robotic assistance, including recent reports of transoral robotic-assisted thyroidectomy. These methods are clearly feasible; however, they do require further study to determine whether there are advantages over the more traditional open approach.

Complications of Thyroidectomy

General complications after thyroid surgery include bleeding and wound complications such as infection and keloid formation. Specific complications include injury to the recurrent laryngeal nerve (<1%), or external branch of the superior laryngeal nerve, temporary hypocalcemia (1.6%-15%), permanent hypocalcemia (<2% for total thyroidectomy), and injury to surrounding structures such as the esophagus, major vessels (carotid artery, internal jugular vein), and the cervical sympathetic trunk.

THE PARATHYROID GLAND

EMBRYOLOGY AND ANATOMY

Around the fourth week of gestation, the human embryo forms five pairs of endoderm-lined pharyngeal pouches. The inferior parathyroid glands are derived from the third branchial pouch (with the thymus), whereas the superior glands arise from the fourth branchial pouch. Most individuals have four parathyroid glands that are found as paired structures on the posterior aspect of the thyroid gland. Supernumerary glands occur in up to 22% of people. Fewer than four glands have also been reported in 3% to 5% of individuals. Approximately 85% of parathyroid glands are found within 1 cm of the point of intersection of the recurrent laryngeal nerve and the inferior thyroid artery. The superior parathyroid glands are usually dorsal to the nerve, whereas the inferior glands are usually ventral to it. The glands may also be found in several ectopic locations. Because the inferior glands migrate farther, they are more often situated in ectopic locations such as the thyrothymic ligament, thymus, carotid sheath, and anterior mediastinum. Ectopically located superior glands can be located in the tracheoesophageal groove, carotid sheath, and posterior mediastinum. Intrathyroidal glands have been reported in 0.5% to 3% of individuals and may be superior or inferior. Normal parathyroid glands are generally embedded in fat, appear golden-yellow to light-brown in color, and weigh approximately 40 to 50 mg. The blood supply to the parathyroid glands is primarily via the inferior thyroid arteries, but the superior thyroid arteries may also supply both the upper and the lower glands.

INDICATIONS FOR SURGERY

PRIMARY HYPERPARATHYROIDISM

This disorder occurs in 1:500 women and 1:2000 men. It is the most common cause of hypercalcemia in the outpatient population, and along with malignancy-associated hypercalcemia accounts for the majority (>90%) of cases of hypercalcemia. Although most cases of PHPT are sporadic, the disorder may also be inherited as
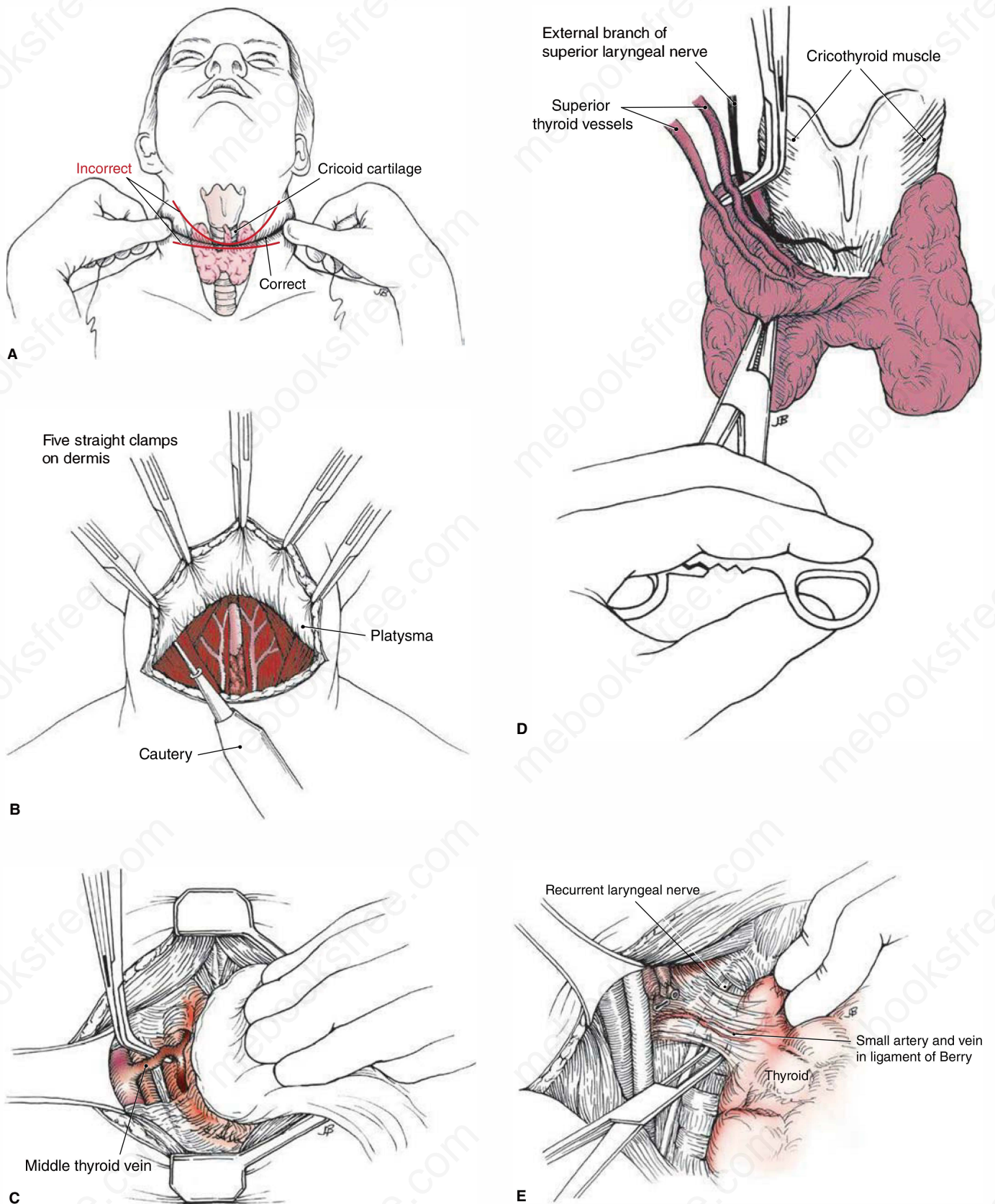
FIGURE 26–2 Conduct of thyroidectomy. **A.** Placement of incision. **B.** Raising subplatysmal flaps. **C.** Dissection of the middle thyroid vein. **D.** Dissection of superior pole vessels. **E.** Dissection at the ligament of Berry. (Reproduced with permission from Lal G, Clark OH. Thyroid, parathyroid, and adrenal. In: *Schwartz's Principles of Surgery*. 8th ed. McGraw-Hill; 2004.)
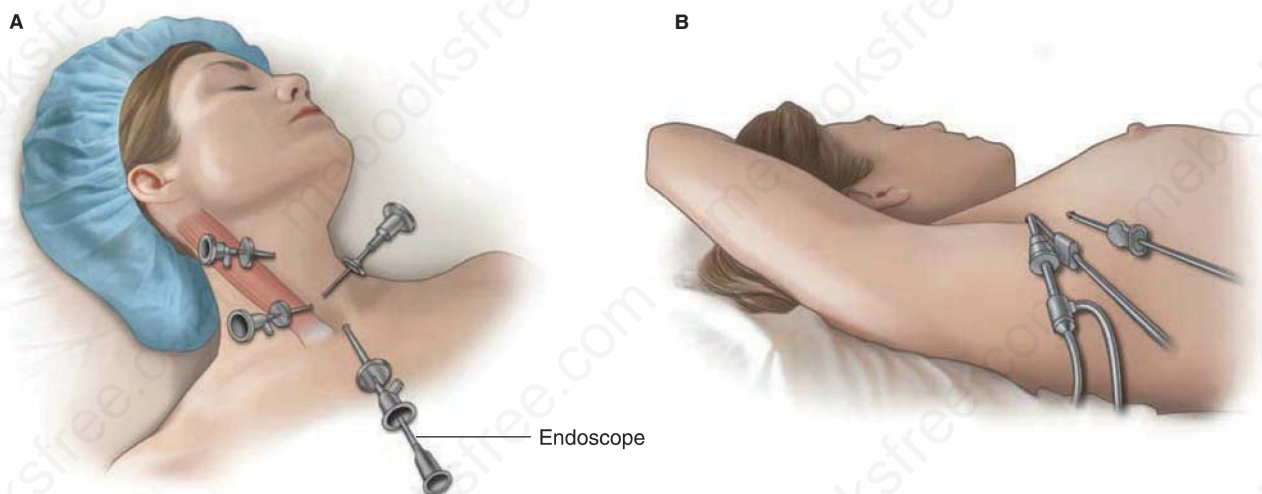
FIGURE 26–3 Endoscopic thyroidectomy can be performed via **(A)** cervical or **(B)** axillary incisions.

a component of MEN 1, MEN 2A, isolated familial hyperparathyroidism, and familial hyperparathyroidism with jaw tumor syndrome. PHPT is the earliest and most common manifestation of MEN 1, and 80% to 100% of patients develop hypercalcemia by age 40. MEN 1 is characterized by germline mutations in the *MENIN* gene located on chromosome 11. In contrast, hyperparathyroidism develops in only 20% of individuals with MEN 2A and is generally less severe. MEN 2A is characterized by inherited mutations in the *RET* proto-oncogene located on chromosome 10. In sporadic cases, hyperparathyroidism results from a single enlarged gland (adenoma) in 85%, multiple enlarged glands (hyperplasia) in 11%, double adenomas in 3%, and parathyroid carcinoma in 1% of patients. Parathyroid carcinoma should be suspected if patients present with a short history, profound hypercalcemia, and a palpable parathyroid gland. In heritable disorders, hyperparathyroidism is more frequently associated with multiple abnormal parathyroid glands and a higher risk of persistent or recurrent disease. Familial hyperparathyroidism with jaw tumor syndrome is associated with an increased risk of parathyroid cancer and results from germline mutations in the *HRPT2* (CDC73 or parafibromin) tumor suppressor gene on chromosome 1.

The classic symptoms of PHPT include painful bones, renal stones, abdominal groans, psychic moans, and fatigue overtones. Other symptoms such as polyuria, nocturia, polydipsia, constipation, and musculoskeletal aches and pains may be present. The disorder may also be associated with hypertension, osteopenia, osteoporosis, nephrolithiasis, gout, pseudogout, peptic ulcer disease, and pancreatitis.

Diagnostic Tests

Other causes of hypercalcemia can generally be excluded by a careful history and physical examination. Laboratory tests that are helpful in making the diagnosis have been discussed in Chapter 8. PHPT is characterized by hypercalcemia (95%), hypophosphatemia (50%), hyperchloremia (30%), elevated 24-hour urinary

calcium, and an increased or inappropriately elevated intact PTH. An elevated alkaline phosphatase level suggests bone disease (osteitis fibrosa cystica). The 24-hour urinary calcium levels may be normal or elevated. Usually, 24-hour urinary studies are obtained unless the patient has previously had documented normocalcemia, to rule out familial hypocaluric hypercalcemia (FHH) type 1, a disorder caused by inherited mutations in the *CASR* gene and characterized by low (often <50 mg) 24-hour urinary calcium excretion. Furthermore, the serum calcium to creatinine clearance ratio is less than 0.01 in patients with FHH type 1 and generally more than 0.02 in patients with PHPT. Two other types of FHH have been described due to mutations in G-alpha 11 (type 2) and adapter protein-2 subunit 1 (AP2S1; type 3) (see Chapter 8).

Parathyroid tumors may be localized using noninvasive studies such as Tc-99m sestamibi (Figure 26–4A) and ultrasound scanning (Figure 26–4B). MRI and CT scans and invasive localizing studies, including highly selective venous catheterization for PTH, and FNAB of suspected parathyroid masses and arteriograms are generally only used in cases of recurrent or persistent hyperparathyroidism. The utility of noninvasive studies prior to initial neck exploration is controversial; however, most surgeons agree that they are indispensable for recurrent or persistent hyperparathyroidism and for using a minimally invasive approach. 99m-Technetium-labeled sestamibi has more than 80% sensitivity for the detection of hypercellular parathyroid glands, and when it is used in conjunction with SPECT, it has particular utility in identifying ectopic tumors. Neck ultrasound is often complementary and has a sensitivity of 75% in experienced centers. More recently, 4-D CT scans have shown tremendous utility in localizing parathyroid tumors with excellent anatomic detail, in addition to providing functional information. It is important to note that most localizing studies are less sensitive in the setting of multiglandular disease. Since the first use of intraoperative PTH in 1993, it has become widely used to assess the adequacy of parathyroid resection. Although preoperative localization studies and a focused surgical approach are being increasingly used,
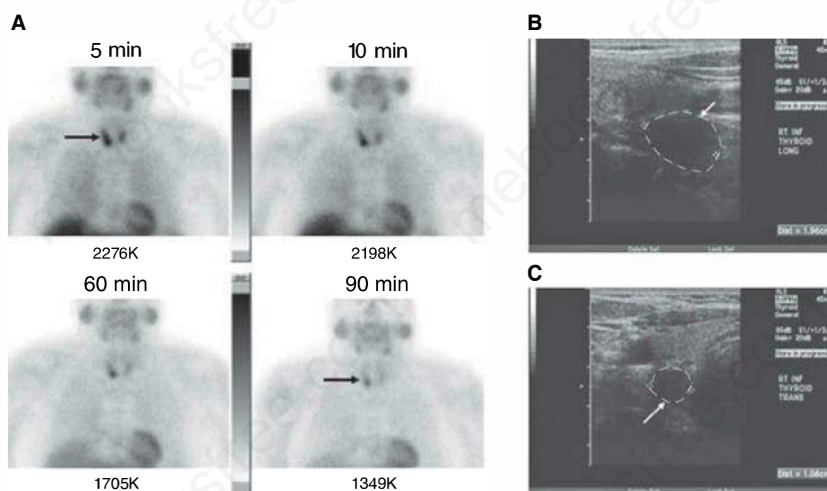
FIGURE 26–4 Parathyroid localization studies. **A.** Sestamibi scan with early (5-10 minutes) and delayed (60-90 minutes) images showing persistent uptake inferior to the right thyroid lobe (black arrow), suggesting a hypercellular parathyroid gland in this location. **B.** Neck ultrasound (longitudinal view). **C.** Neck ultrasound (transverse view) also showing a probable right inferior parathyroid adenoma (white arrow), which is hypoechoic compared to thyroid tissue.

long-term outcome and cost-effectiveness studies will be necessary before recommending their routine use. However, they should be obtained if a focused parathyroidectomy is planned.

Surgical Management

A. Rationale and guidelines for parathyroidectomy

Patients with classic symptoms and metabolic complications related to PHPT are generally treated by parathyroidectomy. Whether asymptomatic patients should undergo parathyroidectomy is controversial. One aspect of the controversy is the fact that there is no consensus on what constitutes an asymptomatic patient. At the 1990 NIH Consensus Conference, asymptomatic PHPT was defined as, "the absence of common symptoms and signs of PHPT, including no bone, renal, gastrointestinal or neuromuscular disorders." Nonoperative management was recommended in this group of patients with calcium levelsless than or equal to 12 mg/dL, based on several observational studies that suggested stability with respect to serum calcium, kidney stones, bone loss, and renal function with time. Parathyroidectomy was recommended for asymptomatic patients less than 50 years of age and in certain other patients.

A second NIH Conference was held in 2002 to reevaluate the criteria for parathyroidectomy in asymptomatic patients. The guidelines issued were similar to the previous recommendations except for the following changes. Parathyroidectomy was recommended in the following cases: for milder elevations in serum calcium (ie, if the serum calcium was 1 mg/dL above reported normal range) and if the bone mineral density at any site (forearm, spine, or hip) was more than 2.5 standard deviations below that of gender-matched peak bone mineral density (T score, rather than Z score, of –2.5). The panel recommended caution in using neuropsychological abnormalities, cardiovascular disease, menopause, gastrointestinal symptoms, and serum or urine indices of increased bone turnover as sole indications for parathyroidectomy. These

recommendations were revised again in 2008, with elevated urinary calcium excretion (in the absence of nephrolithiasis) no longer being a criterion for surgery. In the most recent 2014 revision of the guidelines, urine calcium has become a consideration again and parathyroidectomy is indicated for urine calcium more than 400 mg/d in the presence of increased stone risk by biochemical stone risk analysis. The significant changes from the previous guidelines pertain to the fact that (a) patients with nephrolithiasis or nephrocalcinosis by X-ray, ultrasound or CT scan and (b) those with vertebral fracture by X-ray, CT, MRI, or vertebral fracture assessment (VFA) are also candidates for parathyroidectomy. Recommendations from all these conferences are summarized in Table 26–6.

However, other investigations suggest that true asymptomatic hyperparathyroidism may be rare. Patients without classic symptoms often present with weakness, fatigue, polydipsia, polyuria, nocturia, bone and joint pain, constipation, and other symptoms. These patients also score lower than healthy controls when assessed by multidimensional health assessment tools such as the SF-36 and other more specific questionnaires. Furthermore, parathyroidectomy not only often improves the classic, but also the nonspecific symptoms described above in most patients. In addition, parathyroidectomy has been associated with improved survival (in both asymptomatic and symptomatic patients), is more cost-effective than lifelong follow-up, and is successful in 95% of patients with a minimal (<1%) complication rate. Although bisphosphonates and calcimimetics show promise in the treatment of patients with PHPT, further studies are needed prior to recommending their routine use in this setting. Therefore, most clinicians recommend parathyroidectomy (by surgeons experienced in the procedure) for patients with PHPT, except in those patients where operative risks are prohibitively high. This thought process is also reflected in the 2008 Workshop guidelines and reiterated in the most recent 2014 guidelines on the management of asymptomatic primary hyperparathyroidism.

TABLE 26–6 NIH Consensus Conference and International Workshop indications for surgery in patients with asymptomatic primary hyperparathyroidism.

Criterion	1990	2002	2008	2014
Serum calcium	>12 mg/dL	>1 mg/dL above upper limit of normal	>1 mg/dL above upper limit of normal	>1 mg/dL above upper limit of normal
Urinary calcium	>400 mg/24 h	>400 mg/24 h	Not indicated	>400 mg/24 h
Creatinine clearance	<30% of normal age-matched controls in the absence of another identifiable cause	Reduced by 30% of normal	Reduced to <60 mL/min	Reduced to <60 mL/min
Bone mineral density	<2 SD below age- and sex-matched normal value (Z-score)	<2.5 SD below peak bone mass (T-score) at any site	T-score <−2.5 at any site and/or previous fracture	T-score <−2.5 at any site and/or previous fracture
Age	<50 y	<50 y	<50 y	<50 y

B. Conduct of parathyroidectomy Issues to consider before deciding on the optimal treatment approach are whether one is dealing with sporadic or familial disease and whether it is the initial operation or reexploration. The presence of concurrent thyroid pathology needing surgical treatment must also be considered. The positioning, incision, and dissection are as described previously for thyroidectomy. A bloodless field is crucial to enable identification of parathyroid glands. The middle thyroid veins are ligated and divided to allow medial retraction of the thyroid lobe. The space between the carotid sheath and thyroid is then opened by sharp and blunt dissection, from the level of the cricoid to the thymus, and the recurrent laryngeal nerve is identified. Most parathyroid glands are found within 1 cm of the junction of the nerve and inferior thyroid artery—the superior glands lie dorsal and the lower glands more ventral or anterior to the nerve. Because parathyroid glands are often embedded in fat, any fat lobule at typical parathyroid locations should be explored.

C. Bilateral versus unilateral approach Traditionally, all four parathyroid glands have been explored without preoperative localizing studies, with a 95% success rate in the hands of an experienced surgeon. Investigators in favor of unilateral exploration have suggested decreased recurrent laryngeal nerve injury, less postoperative hypocalcemia, shorter operative time, and early discharge as advantages of this approach. Potential disadvantages of unilateral exploration include the risk of missing multiple abnormal or ectopic parathyroid glands.

More surgeons today are using a unilateral or focal approach directed by localizing studies with or without intraoperative gamma probe or intraoperative PTH measurements. We have found that when two localizing studies—Tc-99m sestamibi and ultrasound—identify the same solitary parathyroid gland in patients with sporadic PHPT, this was the only abnormal parathyroid gland in 95% of patients. In patients with sporadic PHPT and positive preoperative localization studies, our current approach is to recommend focused parathyroidectomy via a 2.5-cm incision with intraoperative PTH measurement. If the PTH level falls more than 50% within 10 minutes after removal of an abnormal parathyroid gland, the procedure is terminated. If the level does not fall, a PTH measurement is repeated in another 10 minutes,

because it may fall slowly in some patients. A bilateral approach is needed in patients with known familial disease, with negative or equivocal localization tests, with secondary or tertiary HPT, or a history of lithium treatment. If parathyroid cancer is suspected intraoperatively, it is resected with the ipsilateral thyroid lobe and regional lymph nodes.

D. Subtotal parathyroidectomy versus total parathyroidectomy with autotransplantation Either subtotal (generally resection of 3½ glands) or total parathyroidectomy with primary parathyroid autotransplantation may be performed in patients with multiglandular disease. In the latter technique, 12 to 20 small fragments (1 × 1 mm) of parathyroid tissue are placed into multiple pockets in the forearm muscle of the nondominant arm. The site is marked with silk sutures and is easily accessible under local anesthesia should the patient develop recurrent hypercalcemia. The recurrence rates for hyperparathyroidism appear to be similar after subtotal or total parathyroidectomy. However, a 5% failure rate has been reported after autotransplantation of parathyroid tissue. Therefore, subtotal parathyroidectomy is preferred. The most normal gland is biopsied first, leaving a 50-mg remnant (the size of a normal parathyroid gland); if it appears viable, the remaining glands are excised. Patients undergoing either of these procedures should have tissue cryopreserved, if possible. Parathyroid glands should not be routinely biopsied at exploration but rather biopsy should be used to confirm parathyroid tissue or help determine if the gland is normal or abnormal. As with thyroid surgery, video-assisted and endoscopic axillary approaches are feasible, but clear benefits compared with the open approach await long-term multicenter studies. The area where endoscopic approaches has shown most promise are in the resection of ectopic mediastinal tumors where thoracoscopic parathyroidectomy is an alternative to sternotomy.

NORMOCALCEMIC PRIMARY HYPERPARATHYROIDISM

This entity is being increasingly recognized and is characterized by the presence of elevated PTH levels with persistently normal serum calcium levels, including ionized serum calcium levels. The diagnosis is established after excluding causes of secondary PTH

elevation such as vitamin D deficiency, osteomalacia, malabsorption, hypercalciuria (renal leak), and renal insufficiency. Medications such as hydrochlorthiazide and lithium should be discontinued. Patients may present with classical features of PHPT including a history of kidney stones (14%), fragility fractures (11%) and osteoporosis (57%). While some symptomatic patients may benefit from parathyroidectomy, close follow-up is also an option due to the lack of long-term outcome data. If the patient develops hyeprcalcemia, the usual parathyroidectomy guidelines may be applied in management decisions.

PERSISTENT AND RECURRENT PRIMARY HYPERPARATHYROIDISM

Persistent hypercalcemia occurs after about 5% to 10% of explorations. Recurrent hypercalcemia is rare except in patients with familial disease and occurs after an intervening period (>6 months) of normocalcemia. The most common reasons for persistent or recurrent hyperparathyroidism are a missed gland in normal or ectopic location, unrecognized hyperplasia, supernumerary glands, subtotal tumor resection, parathyroid carcinoma, or parathyromatosis (usually due to implantation of cells due to spillage at the initial procedure). The most common sites of ectopic glands in patients with recurrent or persistent PHPT (Figure 26–5) are
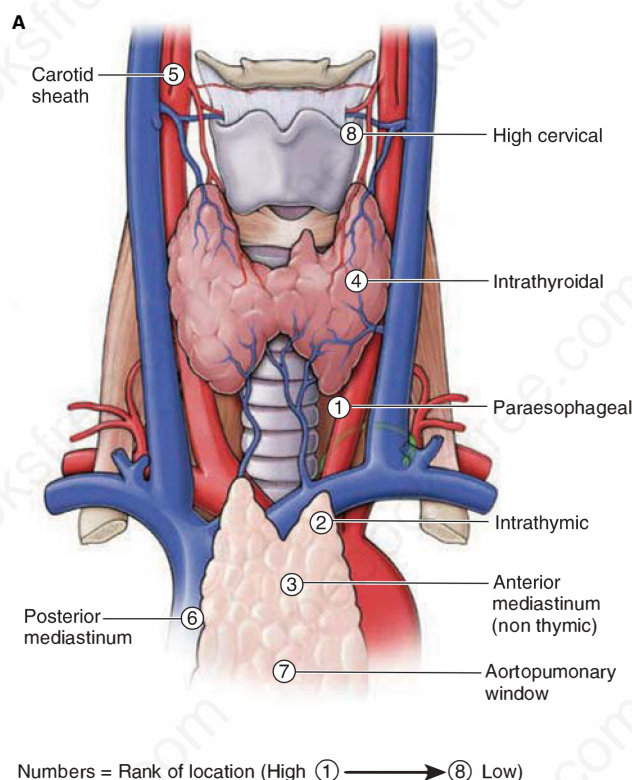
FIGURE 26–5 Ectopic locations of parathyroid tumors found at reoperation after an initial failed neck exploration. The numbers indicate the frequency with which missing glands are found in that location with 1 being most common and 8 being least common.

paraesophageal (28%), mediastinal (26%), intrathymic (24%), intrathyroidal (11%), carotid sheath (9%), and undescended (high cervical 2%). The management of these patients involves confirmation of the diagnosis—in particular, exclusion of FHH; review of original operative notes and pathology reports; and localization studies, which are absolutely essential in this group of patients. Preoperative evaluation of the vocal cords is routinely performed. The issue of reexploration in mildly symptomatic patients is controversial. The neck is usually reexplored first, and median sternotomy may be needed in 1% to 2% of patients. Cryopreservation should be performed routinely and autotransplantation selectively. Reexploration by an experienced surgeon is successful in more than 90% of patients, but reoperative parathyroidectomy is generally associated with more complications than an initial operation.

SECONDARY HYPERPARATHYROIDISM

This disorder occurs most often in individuals with end-stage renal failure but may also occur in other conditions resulting in hypocalcemia, such as vitamin D deficiency, idiopathic hypercalciuria, and long-term lithium therapy.

Indications for surgery include serum calcium and phosphate product more than 70, renal osteodystrophy with or without bone pain, severe pruritus, ectopic soft tissue calcifications and tumoral calcinosis, serum calcium more than 11 mg/dL with markedly elevated PTH, and calciphylaxis.

These patients should undergo dialysis the day prior to surgery to correct electrolyte (specifically potassium) abnormalities. Patients with this disorder require bilateral exploration and either subtotal parathyroidectomy, leaving a 50 to 60 mg histologically confirmed, well-vascularized remnant of hyperplastic parathyroid distant from the recurrent laryngeal nerve; or total parathyroidectomy with autotransplantation of a similar amount of tissue. Upper thymectomy is usually performed in these patients, because up to 15% of patients have a fifth hyperplastic gland.

SPECIAL CONSIDERATION: FAMILIAL HYPERPARATHYROIDISM

Patients from MEN 1 families should have PHPT treated before coexisting gastrinoma. All glands should be identified, ectopic glands should be sought in the neck and upper mediastinum, and bilateral cervical thymectomy should be routinely performed due to the occurrence of thymic carcinoids. A subtotal parathyroidectomy (preferred by authors) or total parathyroidectomy, as described above, can be performed. MEN 2A patients should undergo screening for the presence of a pheochromocytoma and *RET* mutation prior to thyroid or parathyroid surgery. Because the parathyroids are at risk during thyroidectomy and central neck node dissection, and hyperparathyroidism is less virulent in these patients, only obviously enlarged glands should be removed. Normal-appearing glands should be marked and removed abnormal parathyroid tissue cryopreserved.

COMPLICATIONS OF PARATHYROID SURGERY

General complications are similar to those associated with thyroidectomy. Specific complications include recurrent laryngeal nerve injury, hypomagnesemia, and hypocalcemia. The latter may arise due to suppressed function of the remaining glands after removal of an adenoma; injury to the parathyroid remnant; or "bone hunger," which is the influx of calcium and phosphorus into previously demineralized bones. Bone hunger can be predicted to occur based on the severity of bone disease present prior to surgery and at times by the elevated preoperative alkaline phosphatase level and is generally more severe in individuals with uremic secondary hyperparathyroidism. Hypocalcemia can be treated with oral calcium and vitamin D supplementation (calcitriol, 0.25-0.5 μg twice daily). Intravenous calcium is seldom required but may be necessary in severely symptomatic patients.

THE ADRENAL (SUPRARENAL) GLAND

EMBRYOLOGY AND ANATOMY

The adrenals are paired structures located superior to the kidneys. The adrenal is divided into an outer cortex and an inner medulla. The cortex originates from mesodermal tissue near the gonads on the adrenogenital ridge at approximately the fifth week of gestation. Adrenocortical tissue can thus be found in the ovaries, spermatic cord, and testes. The adrenal medulla originates from the neural crest, which is ectodermal in origin.

Each adrenal is supplied by three sets of arteries: the superior adrenal (from the inferior phrenic artery), the middle adrenal (from the aorta), and the inferior adrenal (from the renal artery). These vessels branch into as many as 50 arterioles. The left adrenal vein empties into the ipsilateral renal vein, whereas the right adrenal vein drains into the inferior vena cava (see Chapter 9).

INDICATIONS FOR SURGERY

PRIMARY HYPERALDOSTERONISM

This disorder accounts for 1% of hypertensive patients and results from an adrenal adenoma (Conn syndrome, 75%), adrenal hyperplasia (24%), or adrenocortical cancer (1%). Glucocorticoid-suppressible hyperaldosteronism is another rare cause of hyperaldosteronism. Patients with primary hyperaldosteronism typically present with hypertension (longstanding and difficult to control despite multiple medications), muscle weakness, polydipsia, polyuria, headaches, and fatigue.

Diagnostic Tests

The diagnosis must be suspected in any hypertensive patient with spontaneous hypokalemia or hypokalemia while on diuretic therapy (despite potassium supplementation), although hypokalemia is not required for the diagnosis. Other abnormalities include hypernatremia, alkalosis, and hypochloremia. Elevated plasma and urine aldosterone levels with suppressed plasma renin activity levels confirm the diagnosis. Plasma aldosterone concentration to plasma renin activity ratios of more than 1:25 to 30 are highly suggestive of the diagnosis. The diagnosis can be confirmed by documenting failure to suppress plasma aldosterone with oral or intravenous salt loading. Once the biochemical diagnosis is confirmed, it is important to distinguish between patients with a unilateral adenoma from those with hyperplasia because unilateral adrenalectomy is virtually always curative for the former (see Chapter 10 for details). Tumors can be localized with CT scans, MRI, iodocholesterol scans, or selective venous catheterization for aldosterone and cortisol. A unilateral 0.5- to 2-cm adrenal tumor with a normal-appearing contralateral gland essentially confirms an aldosteronoma if the biochemical and clinical profiles are appropriate.

Surgical Management

Patients are prepared for surgery with potassium repletion, sodium restriction, and treatment with spironolactone or eplerenone (aldosterone antagonists). Amiloride or triamterene (potassium-sparing diuretics) as well as other effective antihypertensive medications, including calcium channel blockers and angiotensin-converting enzyme inhibitors, can also be used. Patients with a unilateral hyperfunctioning tumor are best managed with a laparoscopic adrenalectomy. Postoperatively, patients may require saline infusions, fludrocortisone (50-100 μg/d), or, rarely, glucocorticoids if they develop Addison disease. Adrenalectomy improves blood pressure in approximately 80% of patients and leads to resolution of hypokalemia in most. Only 20% to 30% of patients with bilateral hyperplasia benefit from adrenalectomy. Surgical candidates can be selected based on bilateral adrenal vein sampling for aldosterone and cortisol. A unilateral gradient suggests a potential benefit from surgical intervention. The remaining patients are best managed by medical therapy with spironolactone or eplerenone, triamterene, or amiloride.

HYPERCORTISOLISM

Cushing syndrome refers to a complex of symptoms and physical signs resulting from cortisol hypersecretion, independent of etiology. Causes of Cushing syndrome include ACTH–secreting pituitary tumor leading to bilateral adrenal hyperplasia (Cushing disease 70%); primary adrenal adenoma, hyperplasia, or carcinoma (20%); ectopic ACTH-secreting tumors (small cell lung, pancreatic, thyroid, thymic, and other cancers—10%); or the exogenous administration of steroids. Patients present with weight gain, muscular weakness, polyuria, emotional lability, moon facies, acne, hirsutism, central obesity, hypertension, diabetes mellitus, and virilization (see Chapters 4 and 9).

Diagnostic Tests

Diagnostic tests in patients with Cushing syndrome are directed toward confirming the diagnosis and determining the etiology of

the disorder. A low-dose (1-mg) overnight dexamethasone suppression test is usually performed. Patients with Cushing syndrome do not generally suppress cortisol levels to less than 3 ug/dL, although this threshold is controversial. In patients with a negative test but high clinical suspicion, the classic dexamethasone suppression test (2 mg over 48 hours) or a 24-hour urinary free cortisol measurement is performed. Salivary cortisol measurements are increasingly being used. Once hypercortisolism is confirmed, the plasma ACTH level, if suppressed, indicates a primary adrenal etiology for the disorder. High-dose (8-mg) dexamethasone suppression tests may be performed to delineate the source of increased cortisol secretion (see Chapter 9 for details). Adrenal tumors are localized using CT, MRI, or iodocholesterol scans and pituitary tumors by MRI scanning and selective venous catheterization of petrosal veins after corticotropin-releasing factor stimulation.

Surgical Management

Cushing disease is treated with transsphenoidal hypophysectomy and microsurgical excision of the pituitary adenoma. Irradiation may also be used, but the response is delayed, and the treatment often results in panhypopituitarism. Unilateral laparoscopic adrenalectomy is the treatment of choice in patients with adrenal adenomas, whereas open adrenalectomy is recommended for carcinomas. Bilateral laparoscopic adrenalectomy is used to treat patients with Cushing disease who fail to respond to radiation or hypophysectomy and selected patients with Cushing syndrome secondary to ectopic ACTH production. Symptoms of Cushing syndrome can be treated with medications that inhibit steroid production (ketoconazole, metyrapone, aminoglutethimide).

Preoperatively, electrolyte abnormalities are corrected, and all patients are given exogenous steroids (hydrocortisone, 100 mg intravenously every 8 hours). After unilateral adrenalectomy, steroids are tapered off over months. After bilateral surgery, lifelong treatment with hydrocortisone is necessary. Fludrocortisone is sometimes needed, and steroid supplementation must be increased in situations of stress. The prognosis after adrenalectomy for adenoma is excellent. After bilateral adrenalectomy, approximately 20% of patients develop Nelson syndrome (hyperpigmentation, headaches, exophthalmos, and blindness) from continuing growth of the pituitary tumor.

ADRENAL CORTICAL CARCINOMA

This rare neoplasm is slightly more common in women than in men and has a bimodal age distribution, occurring more frequently in children under 5 years of age and in adults in their forties and fifties. Approximately 50% of these tumors are nonfunctioning. The remaining secrete cortisol (30%), androgens (20%), estrogens (10%), aldosterone (2%), or multiple hormones (35%). Most adrenocortical cancers are sporadic, but they can also occur in the setting of the Li-Fraumeni syndrome (*p53*) and MEN 1 (*MENIN*). Adrenocortical cancers are often characterized by the rapid onset of Cushing syndrome with virilizing features. Nonfunctioning

tumors usually present with an enlarging abdominal mass and abdominal pain or, less commonly, with weight loss, hematuria, varicocele, and dyspnea.

Diagnosis

The biochemical workup of a unilateral adrenal mass is outlined in the section on adrenal incidentaloma (see Chapter 9). CT or MRI scans are commonly employed to evaluate the size and invasiveness of the mass. FDG-PET scans may have some utility in predicting malignancy. The size of the mass remains the single most reliable indicator of malignancy. Carcinomas are more likely to be present in lesions over 6 cm in diameter. Studies based on SEER registry data show that the sensitivity, specificity, and likelihood ratio of tumor size in predicting malignancy as 96%, 51%, and 2, respectively, for tumors larger than or equal to 4 cm; and 90%, 78%, and 4.1, respectively, for tumors larger than or equal to 6 cm. Other features suggesting malignancy on CT include irregular shape and margins, heterogeneity, and hemorrhage. On MRI, carcinomas have amoderate signal intensity on T2-weighted image (adrenal tumor to liver ratio 1.2-2.8). FNAB is usually recommended for patients with an isolated adrenal mass and a history of carcinoma of the lung, breast, stomach, kidney, colon, melanoma, or lymphoma and in those with symptoms and signs of underlying malignancy. FNAB is generally not recommended when a primary adrenal carcinoma is suspected as the diagnosis cannot be reliably made via this test and has the potential to seed the biopsy tract. Care must be taken to exclude pheochromocytoma prior to any biopsy to avoid precipitating a hypertensive crisis (see Chapters 9 and 11).

Surgical Treatment

Complete surgical excision offers the only chance of prolonged survival or cure. Transabdominal adrenalectomy with en bloc excision of contiguously involved structures (liver, kidney, spleen, or pancreas) is usually recommended. A thoracoabdominal approach may also be used for large (>10 cm) right-sided tumors. Laparoscopic adrenalectomy is not usually recommended for suspected adrenal cancers. The adrenolytic agent mitotane and other antitumor drugs such as etoposide, cisplatin, and doxorubicin have also been used with partial success for metastatic tumors. Mitotane has also been associated with an improved recurrence-free survival in the adjuvant setting. Five-year actuarial survival rates of 32% to 48% have been reported in patients who underwent complete resection. Features predicting poor survival include tumor size over 12 cm, six or more mitoses per high-power field, intratumoral hemorrhage, local invasion, and distant metastases.

SEX STEROID EXCESS

Virilization and feminization can be caused by many disorders, including congenital adrenal hyperplasia, adrenal adenomas or carcinomas, ovarian or testicular tumors, hypothalamic or pituitary disease, placental disorders, and exogenous sex steroid administration. Several different variants of congenital adrenal hyperplasia occur, each caused by a specific enzyme defect. Adrenal virilization

occurring postnatally usually results from an adenoma or carcinoma. Virilization presents in females with hirsutism, clitoromegaly, alopecia, breast atrophy, and oligomenorrhea. In males, feminizing tumors lead to gynecomastia and testicular atrophy.

Diagnostic Tests

Karyotype analysis to rule out sex chromosome anomalies such as Klinefelter syndrome and determinations of either plasma and or urinary steroid metabolites are needed to establish the diagnosis of a particular enzyme deficiency as discussed in Chapters 9, 12, and 14. The dexamethasone suppression test (2-4 mg/d in divided doses 4 times daily for 7 days) can be used to distinguish congenital adrenal hyperplasia from neoplasia. CT, MRI, and iodocholesterol scans are used to localize the tumors.

Surgical Management

Congenital adrenal hyperplasia is managed medically not surgically. Adrenalectomy is reserved for treatment of adrenal steroid excess caused by an adenoma or carcinoma. The perioperative management is similar to that for patients with Cushing syndrome. Patients with adrenal carcinomas have a poor prognosis.

PHEOCHROMOCYTOMA

This catecholamine-secreting tumor of the adrenal medulla and extra-adrenal chromaffin tissue accounts for 0.1% to 0.2% of all patients with hypertension. It is often called the "10% tumor" because 10% are bilateral, 10% are malignant, 10% occur in children, 10% are extra-adrenal, and 10% are familial (occurring in association with MEN 2A, MEN 2B, von Hippel-Lindau syndrome, neurofibromatosis, other neurocutaneous syndromes, and the familial paraganglioma and pheochromocytoma syndromes caused by mutations in the succinyl dehydrogenase family of genes [*SDHB, SDHC and SDHD*]). Mutations in SDHA and SDH5 have also been identified. The mTORC1 signaling pathway and MAX (myc-associated factor X) also appear to be involved (see Chapter 11).

Headache, palpitations, and diaphoresis constitute the classic triad of pheochromocytoma. Nonspecific symptoms include anxiety, tremulousness, severe headaches, paresthesias, flushing, chest pain, shortness of breath, abdominal pain, nausea, and vomiting as well as others. The most common clinical sign is hypertension, which may be sustained or episodic.

Diagnostic Tests

These are performed in a nonstressed patient. Twenty-four-hour urine collections are tested for catecholamines (epinephrine, norepinephrine, dopamine) and their metabolites (metanephrines and normetanephrines). Plasma metanephrines, catecholamines, and chromogranin A levels are also used to make the diagnosis. Provocative tests such as glucagon infusion and clonidine suppression are rarely necessary. Radiologic investigations such as CT or

MRI scans, and metaiodobenzylguanidine scans are used to localize the tumors and assess for possible extra-adrenal tumors (see Chapter 11). Of note, extra-adrenal sites lack phenylethanolamine-*N*-methyltransferase, and hence, secrete primarily norepinephrine rather than epinephrine.

Surgical Treatment

Adrenalectomy is the treatment of choice and can safely be performed laparoscopically. Preoperative preparation typically consists of treatment with an alpha-adrenergic blocker such as phenoxybenzamine (10-40 mg 4 times daily; maximum: 300 mg/d) for at least 7 days preoperatively. Beta-blockers such as propranolol (5-40 mg 4 times daily) are added in patients who have persistent tachycardia and arrhythmias but only after adequate alpha blockade has been established in order to avoid the effects of unopposed alpha stimulation (ie, hypertensive crisis and congestive heart failure). Calcium channel blockers such as nicardipine appear to be as effective as alpha and beta blockade preoperatively. Patients should also be volume repleted to avoid intraoperative or postoperative hypotension, which ensues with the loss of vasoconstriction after tumor removal. Nitroprusside is the drug of choice for intraoperative hypertension. After surgery, 95% of patients with paroxysmal hypertension and 65% with sustained hypertension become normotensive. Patients with malignant pheochromocytoma, defined by invasion into surrounding structures or presence of distant metastases, have a poor prognosis. Although risk of malignancy increases with size for all pheochromocytomas, size does not seem to reliably predict malignancy in pheochromocytomas. The risk of malignant tumors is also increased in patients with germline *SDHB* mutations.

ADRENAL INCIDENTALOMA

The term **incidentaloma** is used to denote an adrenal mass discovered during imaging done for other reasons. Adrenal masses have been identified in up to 8% of individuals in autopsy series and in 4.4% of those undergoing abdominal CT scans. The widespread use of ultrasound, CT, and MRI scans over the past two decades has led to an appreciable increase in the number of these lesions identified. Most of these lesions are benign nonfunctioning adenomas. The differential diagnosis of these lesions is summarized in Table 26–7.

Diagnosis

The workup is used to discern which lesions are functional or malignant and thus, warrant adrenalectomy. Asymptomatic patients with obvious cysts, hemorrhage, myelolipomas, or diffuse metastatic disease do not mandate further testing. All other patients should undergo biochemical testing for hormonally active tumors. At a minimum, this includes serum electrolytes; low-dose (1-mg) dexamethasone suppression testing; and a 24-hour urine collection for catecholamines, metanephrines, and 17-ketosteroids; however, the issue of which tests to order is controversial. Confirmatory tests can be performed based on the results of these screening tests. It is

TABLE 26–7 Differential diagnosis of an adrenal incidentaloma.

Benign	Malignant
Adrenal cortex	
Functioning adenoma	Adrenocortical cancer
Nonfunctioning adenoma	
Adrenal medulla	
Pheochromocytoma	Malignant pheochromocytoma
Others	
Cysts	Metastasis
Myelolipomas	
Ganglioneuroma	
Hematoma	

extremely important to identify patients with subclinical Cushing syndrome. Functional tumors and nonfunctional masses over 4 cm (in patients who have an acceptable operative risk) are treated by laparoscopic adrenalectomy, whereas heterogeneous, irregular, or enlarging tumors are treated by open or laparoscopic adrenalectomy. FNAB should be performed only in patients with a history of carcinoma and a suspected isolated adrenal metastasis, or if there is a question as to whether the tumor is of adrenal origin (see Chapter 9 for details). Care must be taken to exclude pheochromocytoma prior to biopsy to avoid precipitating a hypertensive crisis.

Treatment

Laparoscopic adrenalectomy has become the procedure of choice for most of the lesions described above except when malignancy is likely or suspected. Patients with nonfunctioning homogeneous lesions less than 4 cm in diameter should be followed with serial examinations and CT or MRI scans at 3 to 12 months. Adrenalectomy is indicated for any lesion that grows during the observation period.

TECHNIQUE OF ADRENALECTOMY

There are no randomized controlled trials comparing laparoscopic and open adrenalectomies. However, several retrospective studies have shown that the laparoscopic technique is safe and associated with less postoperative pain, shorter mean hospital stay, lower morbidity, and more rapid complete recovery. Laparoscopic adrenalectomy has become the procedure of choice for most adrenal lesions, and the indications are similar to those for open procedures.

Laparoscopic adrenalectomy is contraindicated in patients with a known adrenocortical cancer or coagulopathy. Two approaches have been described: transperitoneal (lateral and anterior) and posterior retroperitoneal. The former provides a conventional view of the anatomy, and the anterior approach allows for bilateral procedures without repositioning the patient. The posterior approach may be preferable in reoperative cases and obese patients but provides a limited working space. The lateral transperitoneal approach is described in the next paragraph.

The patient is placed in the lateral decubitus position, with the table flexed to open the space between the lower rib cage and the iliac crest. Pneumoperitoneum is created by insufflating carbon

dioxide gas and four 10-mm trocars are placed between the midclavicular line medially and the anterior axillary line laterally (Figure 26–6A), approximately two finger-breadths below the costal margin. A 30-degree laparoscope inserted via the midclavicular port is used to aid the dissection, which is carried out via atraumatic instruments inserted via the two most lateral ports. In general, the dissection of the adrenal usually begins superomedially and then proceeds inferiorly in a clockwise fashion (Figures 26–6B and 26–6C). The adrenal gland is dissected from
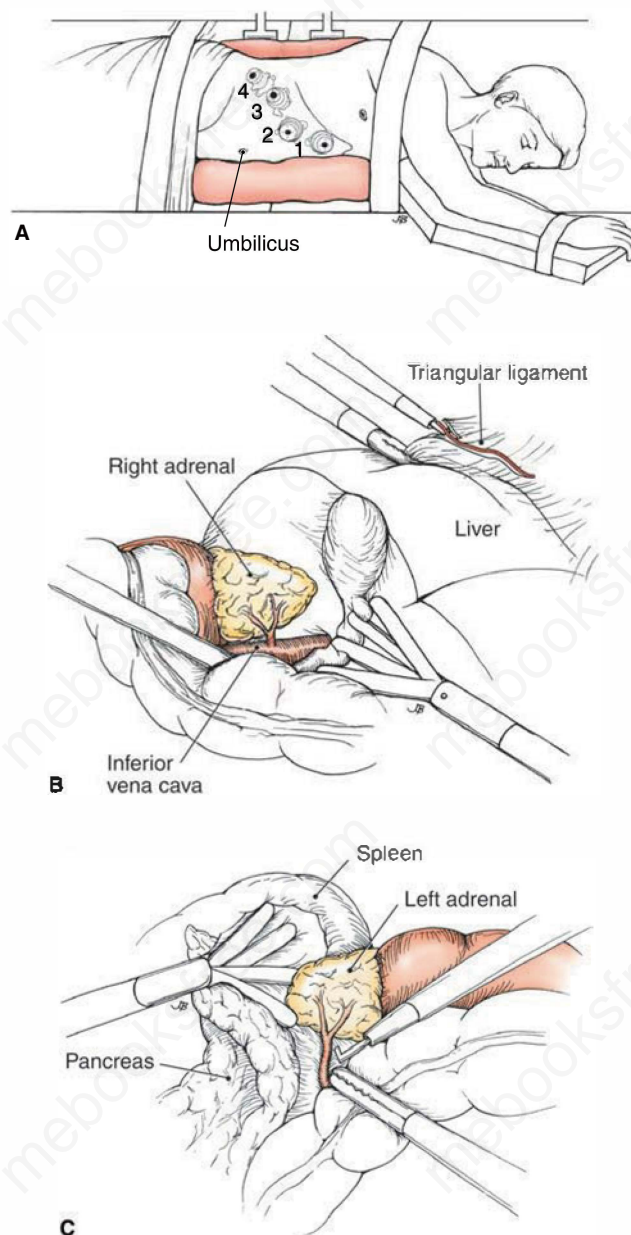
FIGURE 26–6 Technique of laparoscopic adrenalectomy. **A.** Positioning and placement of trocars. **B.** Exposure of the right adrenal is facilitated by division of the triangular ligament. **C.** Dissection and reflection of the spleen and tail of the pancreas aids in identifying the left adrenal. (Reproduced with permission from Brunicardi F, Andersen D, Billiar T, et al. *Schwartz's Principles of Surgery*, 8th ed. New York: McGraw-Hill; 2004.)

surrounding tissue using an electrocautery or an ultrasonic scalpel (Harmonic Ultrasonic Scalpel). These instruments are also useful for control of the small adrenal arteries, but the larger adrenal arteries and adrenal veins need to be clipped or stapled. The adrenal is placed in a specimen (EndoCatch) bag and can be morcellated prior to extraction.

Adrenalectomy can also be performed via the posterior retroperitoneal approach, especially in patients who have had prior abdominal surgery. Another advantage is that bilateral adrenalectomy can be performed without repositioning the patient. However, the working space is limited making vascular control difficult. This technique is being increasingly used for small adenomas but is less suitable for tumors larger than 5 cm.

Open adrenalectomy may be performed by the anterior, posterior, retroperitoneal, or thoracoabdominal approaches. The latter is very useful, particularly for large adrenal cancers.

Complications of Laparoscopic Adrenalectomy

Specific procedure-related complications include trocar site-associated hematoma and subcutaneous emphysema, injury to surrounding organs such as the spleen, pancreas, liver and colon, and bleeding from venous injuries.

THE ENDOCRINE PANCREAS

EMBRYOLOGY AND ANATOMY

The pancreas is a retroperitoneal organ located at the level of L2. It weighs 75 to 100 g, is approximately 15 to 20 cm in length, and is divided into the head and uncinate process, the neck, the body, and the tail. The uncinate process forms part of the head and surrounds the superior mesenteric vessels. The main pancreatic duct (duct of Wirsung) is 2 to 3.5 mm wide, runs in the center of the pancreas, and drains the body, tail, and uncinate process. The lesser duct (duct of Santorini) usually drains the head, communicates with the duct of Wirsung, and drains separately via a minor papilla located 2 cm proximal to the ampulla of Vater. The common bile duct is found posteriorly in the pancreatic head and joins the main pancreatic duct before draining into the ampulla.

The pancreas originates as dorsal and ventral pancreatic buds from the primitive endoderm at about the fifth week of gestation. The former gives rise to the superior head, neck, body, and tail, whereas the later forms the inferior head and the uncinate process. The ventral duct fuses with the dorsal bud to form the duct of Wirsung, and the proximal portion of the dorsal duct forms the duct of Santorini. In 10% of individuals, the ducts fail to communicate, resulting in pancreas divisum, where the entire pancreas is drained by the lesser duct.

INDICATIONS FOR SURGERY

Endocrine pancreatic tumors arise from the islet cells, which are derived from the neural crest. Common functioning tumors are described next.

INSULINOMA

This β-cell–derived neoplasm is the most common pancreatic endocrine tumor. Insulinomas are evenly distributed throughout the pancreas (one-third each in the head, body, and tail). Most patients (90%) have a benign, solitary lesion. Approximately 10% of patients have malignant insulinomas with metastatic disease to the liver and peripancreatic lymph nodes. The insulinoma syndrome is characterized by the Whipple triad, which includes symptoms of hypoglycemia during fasting, serum glucose less than 50 mg/dL, and relief of hypoglycemic symptoms by exogenous glucose administration. Hypoglycemic symptoms may result from neuroglycopenia (confusion, seizures, personality change, coma) or due to a catecholamine surge (tachycardia, diaphoresis, trembling). Other causes of hypoglycemia such as reactive hypoglycemia, adrenal insufficiency, end-stage liver disease, nonpancreatic tumors (mesothelioma, sarcoma, adrenal carcinoma, carcinoid), and surreptitious administration of oral hypoglycemics or insulin should be excluded.

Diagnostic Tests

The diagnosis is made during a 72-hour monitored fast if glucose levels fall to less than 50 mg/dL, while insulin levels are more than 20 uU/mL and the insulin to glucose ratio is more than 0.4 (normal: <0.3). This is considered the gold-standard test. Increased levels of C peptide and proinsulin are also diagnostic, whereas low levels suggest factitious hyperinsulinemia.

Once the diagnosis is confirmed biochemically, noninvasive tests such as double-contrast, fine-cut (5 mm) CT or MRI scans can identify large tumors or liver metastases. Transgastric endoscopic ultrasound is the most successful preoperative localization test (sensitivity 83%-93%). Some pancreatic neuroendocrine tumors also express somatostatin receptors and can be imaged using radiolabeled octreotide. Unfortunately, this is only effective in about 30% of patients with insulinomas (see Chapter 18 for details). PET/CT scan with ^{68}Ga-somatostatin analogs may be helpful. Selective arteriography with calcium stimulation of insulin secretion, monitored through a catheter positioned in the right hepatic vein (Imamura-Doppmann test, sensitivity 88%) is an invasive test that is used selectively when other localization studies fail to identify the tumor or in patients following unsuccessful prior surgery. Glucagon-like peptide-1 (GLP-1) receptor expression is high in most insulinomas and may be promising in localizing small tumors.

Treatment

A. Surgical treatment   Operation is the only curative treatment. Prior to surgery, patients are instructed to take several small frequent meals, and diazoxide administration can be helpful in avoiding hypoglycemic attacks. Other medications commonly used include verapamil, diphenylhydantoin, and glucocorticoids. Somatostatin analogues can also be used if the tumors are positive for somatostatin receptors. Glucose levels are monitored perioperatively. The combination of inspection, palpation, and

intraoperative ultrasound allows detection of nearly all tumors and their relationship to the pancreatic duct. Small (<2 cm) benign tumors in any part of the pancreas that are not intimately associated with the main pancreatic duct are best treated by enucleation. Larger tumors (up to 5 cm) can be enucleated if located in the pancreatic head but are removed by spleen-preserving distal pancreatectomy if located in the tail. Large tumors in the head that appear malignant are usually resected by a Whipple procedure. When a tumor cannot be identified, blind distal resections should generally not be performed. However, a small distal pancreatic resection may be sufficient to rule out nesidioblastosis (β-cell hyperplasia). Resection of peripancreatic and duodenal nodes is advised in patients with probable malignant tumors. Hepatic resection should be considered for cure or palliation in patients with metastatic tumors. In cases of nonresectable disease, debulking may be reasonable to achieve symptomatic control and may also prolong survival. Other modalities such as RFA, cryotherapy, and transarterial chemoembolization (TACE) may also be considered in this setting. Patients with insulinomas in the setting of MEN 1, in contrast to those with sporadic tumors, require distal pancreatectomy and enucleation of tumors from the head of the pancreas. Surgical resection is curative in about 95% of cases.

B. Medical treatment   Diazoxide and verapamil are often used to decrease insulin secretion from insulinoma. Combination chemotherapy (streptozocin, 5-fluorouracil (FU), and doxorubicin) has also been used for unresectable insulinomas.

GASTRINOMA (ZOLLINGER-ELLISON SYNDROME)

This neoplasm leads to abdominal pain, peptic ulceration of the proximal gastrointestinal tract (90%) and diarrhea (50%). The diagnosis should be suspected in patients with recurrent postoperative and postbulbar ulcers, ulcers associated with diarrhea, a family history of ulcer diathesis or MEN 1, and failure to respond to adequate medical therapy. Up to 75% of gastrinomas occur sporadically. In contrast to insulinomas, 60% of gastrinomas are malignant and present with local invasion and metastases. Also, virtually all insulinomas are situated within the pancreas, but up to 70% of gastrinomas are found within the duodenum.

Diagnostic Tests

Fasting serum gastrin levels are generally over 200 pg/mL, and values over 1000 pg/mL are diagnostic of gastrinomas except in hypochlorhydric patients. Elevated gastrin levels may also be found in several other disease states (Table 26–8), and other tests are therefore necessary. Basal gastric acid output over 15 mEq/h (>5 mEq/h in patients with previous vagotomy) or a ratio of basal to maximal acid output more than 0.6 suggests a gastrinoma. An increase of gastrin over 200 pg/mL above the basal level on stimulation with secretin (2 U/kg) confirms the diagnosis. In patients with Zollinger-Ellison syndrome (ZES), serum PTH, calcium, and prolactin levels should be measured to rule out MEN 1.

TABLE 26–8 Causes of hypergastrinemia.

A. Hypergastrinemia associated with increased gastric acid
  1. Gastrinoma, sporadic, or familial (MEN 1)
  2. Antral G cell hyperfunction (rare)
  3. Retained gastric antrum
  4. Short bowel syndrome
  5. Gastric outlet obstruction
  6. Renal failure (acid can be normal)
  7. *Helicobacter pylori* gastritis (acid can be low)
B. Hypergastrinemia associated with little or no gastric acid
  1. Pernicious anemia (achlorhydria)
  2. Chronic atrophic gastritis
  3. Vagotomy
  4. Gastric ulcer associated with hypochlorhydria

Imaging studies used to localize gastrinomas are similar to those used for investigation of insulinomas. However, endoscopic ultrasound and visceral angiograms are less sensitive for gastrinomas, because many are small and not within the pancreas. Most pancreatic gastrinomas are situated to the right of the superior mesenteric vessels. Octreotide scans or positron emission tomography (PET) with Ga-DOTATOC may be useful. Selective pancreatic angiograms utilizing secretin stimulation and hepatic venous sampling for gastrin levels may also be used.

Treatment

A. Surgical treatment   Patients are treated with proton pump inhibitors preoperatively. Most gastrinomas are located in the gastrinoma triangle, which is surrounded by the cystic duct superiorly, the second and third portions of the duodenum inferiorly, and the junction of the neck and body of the pancreas medially. The Whipple procedure is recommended for large (>6 cm) or clinically malignant tumors in the head of the pancreas. Intraoperative ultrasound, endoscopy with duodenal transillumination, and a longitudinal duodenotomy are necessary to identify many tumors. All peripancreatic and periduodenal lymph nodes should be removed, as primary lymph node gastrinomas have been reported. Single hepatic metastases may be resected. Total gastrectomy is rarely indicated today except in noncompliant patients or those refractory to medical therapy when the gastrinoma cannot be identified or completely removed. Patients with MEN 1 often have multiple gastrinomas. The extent and role of surgery in patients with MEN 1/ZES remains controversial. Some groups advise early and aggressive surgeries to prevent liver disease, whereas others advise resection of MEN 1 gastrinomas larger than 2 cm as small tumors are characterized by a good prognosis. In advanced, surgically incurable tumors, cytoreductive measures including resection or RFA/TACE are reasonable if more than 90% of the tumor mass can be removed, although this does not appear to prolong survival. Liver transplantation may be a consideration in young patients with diffuse, unresectable pancreatic metastases. In contrast to patients with MEN 1 who have nearly 100% recurrence rates, about 34% of patients with sporadic gastrinomas remain disease-free at 10 years after surgical treatment. Overall survival rates after surgical treatment are reported to be 94% at 10 years.

B. Medical treatment   Octreotide may be used to decrease gastrin secretion. The combination of doxorubicin, streptozotocin, and 5-FU can be used as systemic chemotherapy. Somatostatin analogues and interferons have not shown much efficacy.

VIPOMA (VERNER-MORRISON) SYNDROME

This disorder is also known as the WDHA (watery diarrhea, hypokalemia, achlorhydria) or pancreatic cholera syndrome. Patients typically present with high-volume diarrhea (>5 L/d), muscular weakness (due to hypokalemia) and lethargy, hyperglycemia, hypercalcemia, and, rarely, cutaneous flushing.

Diagnostic Tests

Because secretion of vasoactive intestinal polypeptide (VIP) can be episodic, multiple fasting levels should be measured. Localizing studies are performed as described earlier for insulinomas and gastrinomas.

Treatment

A. Surgical treatment   Fluid and electrolytes should be aggressively repleted prior to surgery. The diarrhea often responds to octreotide. Most of these tumors are located in the body and tail and hence are best managed by a distal pancreatectomy. Small tumors may be enucleated. If the tumor is not localized, the autonomic chain and adrenals should be examined for extrapancreatic tumors. If no tumor is identified, distal pancreatectomy may be considered. Palliative debulking may be performed for metastatic disease to control symptoms if more than 90% of the tumor mass can be resected. Other ablative therapies such as RFA/TACE or cryotherapy may also be used as cytoreductive measures and liver transplantation may be an option in selected cases.

B. Medical treatment   Somatostatin analogues may provide palliation of symptoms particularly if the tumors are octreotide scan positive. They may also have an effect on tumor growth. Interferon-α may be indicated in tumors not responding to somatostatin analogues. Systemic therapy as described for other tumors may also be used.

GLUCAGONOMA

Patients with this tumor have mild diabetes, stomatitis, anemia, malnutrition, hypoproteinemia, and a characteristic severe dermatitis (necrolytic migratory erythema). The latter is thought to be secondary to the hypoaminoacidemia.

Diagnostic Tests

The clinical presentation and biopsy of the rash are sufficient for the diagnosis. Hyperglycemia, hypoproteinemia, and an elevated fasting glucagon level (>150 pg/mL) confirm the diagnosis.

Localizing and staging tests are performed, as described above for the more common tumors. Most of these tumors are malignant.

Treatment

A. Surgical treatment   Patients need preoperative octreotide, hyperalimentation, and routine deep venous thrombosis prophylaxis. Up to 70% of these tumors present with metastases, and surgery is the only potentially curable treatment. Most tumors are solitary and located in the tail, where they are amenable to distal pancreatectomy. Palliative debulking with distal pancreatectomy and splenectomy (if needed) in the setting of metastatic disease may help refractory symptoms. If splenectomy is planned, appropriate vaccination to reduce the risk of overwhelming post-splenectomy sepsis is recommended preoperatively.

B. Medical treatment   As with other neuroendocrine tumors, octreotide and embolization of hepatic metastases may help control symptoms.

SOMATOSTATINOMA

The somatostatinoma syndrome is characterized by steatorrhea, diabetes, hypochlorhydria, and gallstone disease. Most somatostatinomas are malignant (75% have metastases at presentation) and are located in the pancreatic head. The diagnosis is established by fasting somatostatin levels of 100 pg/mL or more. Somatostatinomas are usually treated with the Whipple procedure (ie, pancreaticoduodenectomy). Fluid and electrolyte abnormalities should be corrected preoperatively.

NONFUNCTIONING PANCREATIC TUMORS

Approximately 33% of patients with pancreatic endocrine neoplasms have no evidence of a defined clinical syndrome and are deemed to have nonfunctioning tumors. However, some of these tumors produce pancreatic polypeptide and chromogranin A, particularly in the setting of MEN 1. These patients usually present with abdominal pain, weight loss, and jaundice—similar to ductal adenocarcinoma of the pancreas. The tumors are most commonly present in the head, neck, and uncinate process. The tumors are localized and staged similar to functional tumors. Octreotide scans are recommended to assess receptor expression status and assess extent of disease. Endoscopic retrograde cholangiopancreatography and percutaneous transhepatic cholangiography are used for the evaluation of jaundice.

Surgical Treatment

Approximately 75% of these tumors are malignant. Surgical resection (pancreaticoduodenectomy or distal pancreatectomy) is the treatment of choice for sporadic tumors, because these tumors are not amenable to enucleation. In the presence of MEN 1, the indications for surgery are controversial. More aggressive groups recommend

tumor enucleation in the head and prophylactic subtotal pancreatectomy, whereas more conservative groups recommend close follow-up and enucleation of lesions larger than 2 cm. Biliary and gastric bypasses may be necessary for palliation. These tumors often grow slowly. Five-year survival rates after resection are 50%. Resectable liver disease should be treated with surgery and or RFA. In patients with unresectable disease, combination chemotherapy (streptozocin and doxorubicin) with RFA/TACE may provide palliation. Poorly differentiated tumors are characterized by aggressive tumor biology, absence of somatostatin receptors, and poor prognosis. Surgical resection is recommended only for resectable primary tumors. Cytoreductive procedures are generally not recommended in these patients.

Novel Therapies

Patients with advanced pancreatic neuroendocrine tumors not amenable to resection may be treated with everolimus (a serine-threonine kinase mTOR inhibitor) and sunitinib (a tyrosine-kinase inhibitor). Both targeted therapies have been associated with improved progression-free survival. Isotope therapy with labeled somatostatin analogs (peptide receptor radionuclide therapy) is also being increasingly used as palliative therapy. Pneumococcal vaccine should be given at least 2 weeks preoperatively in all patients having a splenectomy. Additionally, vaccines against Hemophilus influenza B, meningococcus, and Influenza are also recommended.

Technique of Pancreatic Exploration for Neuroendocrine Tumors

Patients are explored through a midline or bilateral subcostal incision. A detailed description of the Whipple procedure and a distal pancreatectomy is beyond the scope of this chapter, and the reader is referred to surgical atlases. A general description of the necessary maneuvers is presented here. The gastrocolic ligament and the inferior retroperitoneal attachments of the pancreas are divided to permit examination of the body and tail. A Kocher maneuver is performed to facilitate bimanual examination of the head and uncinate process. The duodenum, splenic hilum, small bowel mesentery, gonads (in women), and lymph nodes are assessed for extrapancreatic disease. The liver is examined for metastatic disease. Intraoperative ultrasound facilitates identification of tumors. A laparoscopic approach is also feasible in these tumors, particularly sporadic insulinomas, when the lesion is identified with preoperative localization studies and/or intraoperative ultrasound. Minimally invasive procedures also appear to be safe in the treatment of benign and malignant pancreatic nueoendocrine tumors without compromising oncologic outcomes. However, larger, multicenter studies are needed to evaluate them in relation to traditional open approaches. Liver transplantation may be considered in selected patients with difficult to control hormone oversecretion-related symptoms and low-risk tumors without distant metastases.

Complications of Pancreatic Surgery

The most important complications of pancreatic surgery (tumor enucleation, distal pancreatectomy, or Whipple resection) are pancreatic fistula, pseudocyst, or abscess formation, which may lead to necrotizing retroperitoneal infection, and hemorrhage. Other complications include upper gastrointestinal tract bleeding, marginal ulceration, and biliary fistula formation. The mortality of a Whipple procedure is less than 5% and that of other pancreatic surgical procedures is less than 1%.

REFERENCES

The Thyroid Gland

Bahn RS, Burch HB, Cooper DS, et al. Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. *Endocr Pract*. 2011;17:456. [PMID: 21700562]

Cibas ES, Ali SZ, NCI Thyroid FNA State of the Science Conference. The Bethesda System for Reporting Thyroid Cytopathology. *Am J Clin Pathol*. 2009;132:658. [PMID: 19846805]

Eskander A, Merdad M, Freeman JL, Witterick IJ. Pattern of spread to the lateral neck in metastatic well-differentiated thyroid cancer: a systematic review and meta-analysis. *Thyroid*. 2013;23:583. [PMID: 23148663]

Francis GL, Waguespack SG, Bauer AJ, et al. Management guidelines for children with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2015;25:716. [PMID: 25900731]

Genovese BM, Noureldine SI, Gleeson EM, Tufano RP, Kandil E. What is the best definitive treatment for Graves' disease? A systematic review of the existing literature. *Ann Surg Oncol*. 2013;20:660. [PMID: 22956065]

Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26:1. [PMID: 26462967]

Linos D. Minimally invasive thyroidectomy: a comprehensive appraisal of existing techniques. *Surgery*. 2011;150:17. [PMID: 21596415]

Mayson SE, Yoo DC, Gopalakrishnan G. The evolving use of radioiodine therapy in differentiated thyroid cancer. *Oncology*. 2015;88:247. [PMID: 25503797]

Pisanu A, Porceddu G, Podda M, Cois A, Uccheddu A. Systematic review with meta-analysis of studies comparing intraoperative neuromonitoring of recurrent laryngeal nerves versus visualization alone during thyroidectomy. *J Surg Res*. 2014;188:152. [PMID: 24433869]

Smallridge RC, Ain KB, Asa SL, et al. American Thyroid Association guidelines for management of patients with anaplastic thyroid cancer. *Thyroid*. 2012;22:1104. [PMID: 23130564]

Sturgeon C, Yang A, Elaraj D. Surgical management of lymph node compartments in papillary thyroid cancer. *Surg Oncol Clin N Am*. 2016;25:17. [PMID: 26610772]

Wells SA Jr, Asa SL, Dralle H, et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 2015;25:567. [PMID: 25810047]

The Parathyroid Gland

Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3561. [PMID: 25162665]

Cusano NE, Silverberg SJ, Bilezikian JP. Normocalcemic primary hyperparathyroidism. *J Clin Densitom*. 2013;16:33. [PMID: 23374739]

Eastell R, Brandi ML, Costa AG. Diagnosis of asymptomatic primary hyperparathyroidism: proceedings of the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3570. [PMID: 25162666]

Fang SH, Lal G. Parathyroid cancer. *Endocr Pract*. 2011;17 Suppl 1:36. [PMID: 21454239]

Gracie D, Hussain SS. Use of minimally invasive parathyroidectomy techniques in sporadic primary hyperparathyroidism: systematic review. *J Laryngol Otol*. 2012;126:221. [PMID: 22032618]

Irvin GL 3rd, Solorzano CC, Carneiro DM. Quick intraoperative parathyroid hormone assay: surgical adjunct to allow limited parathyroidectomy, improve success rate, and predict outcome. *World J Surg*. 2004;28:1287. [PMID: 15517474]

Udelsman R, Åkerström G, Biagini C, et al. Surgery for asymptomatic primary hyperparathyroidism: proceedings of the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3595. [PMID: 25162669]

The Adrenal Gland

Broome JT, Gauger P. Surgical techniques for adrenal tumors. *Minerva Endocrinol*. 2009;34:185. [PMID: 19471241]

Chai YJ, Kwon H, Yu HW, et al. Systematic review of surgical approaches for adrenal tumors: lateral transperitoneal versus posterior retroperitoneal and laparoscopic versus robotic adrenalectomy. *Int J Endocrinol*. 2014;2014:918346. [PMID: 25587275]

Conzo G, Tartaglia E, Gambardella C, et al. Minimally invasive approach for adrenal lesions: systematic review of laparoscopic versus retroperitoneoscopic adrenalectomy and assessment of risk factors for complications. *Int J Surg*. 2016;28 Suppl 1:S118. [PMID: 26708860]

Datta J, Roses RE. Surgical Management of adrenocortical carcinoma: an evidence-based approach. *Surg Oncol Clin N Am*. 2016;25:153. [PMID: 26610780]

Fishbein L. Pheochromocytoma and paraganglioma: genetics, diagnosis, and treatment. *Hematol Oncol Clin North Am*. 2016;30:135. [PMID: 26614373]

National Institutes of Health. NIH state-of-the-science statement on management of the clinically inapparent adrenal mass ("incidentaloma"). *NIH Consensus State Sci Statements*. 2002;19:1. [PMID: 14768652]

Rehan M, Raizman JE, Cavalier E, Don-Wauchope AC, Holmes DT. Laboratory challenges in primary aldosteronism screening and diagnosis. *Clin Biochem*. 2015;48:377. [PMID: 25619896]

Rossi GP, Auchus RJ, Brown M, et al. An expert consensus statement on use of adrenal vein sampling for the subtyping of primary aldosteronism. *Hypertension*. 2014;63:151. [PMID: 24218436]

Tang K, Li H, Xia D, et al. Robot-assisted versus laparoscopic adrenalectomy: a systematic review and meta-analysis. *J Laparoendosc Adv Surg Tech A*. 2015;25:187. [PMID: 25763475]

Starker LF, Kunstman JW, Carling T. Subclinical Cushing syndrome: a review. *Surg Clin North Am*. 2014;94:657. [PMID: 24857582]

Walz KM. Extent of adrenalectomy for adrenal neoplasm: cortical sparing (subtotal) versus total adrenalectomy. *Surg Clin North Am*. 2004;84:743. [PMID: 15145232]

The Endocrine Pancreas

Cloyd JM, Poultsides GA. Non-functional neuroendocrine tumors of the pancreas: advances in diagnosis and management. *World J Gastroenterol*. 2015;21:9512. [PMID: 26327759]

Dravecka I, Lazurova I. Nesidioblastosis in adults. *Neoplasma*. 2014;61:252. [PMID: 24645840]

Farley HA, Pommier RF. Treatment of neuroendocrine liver metastases. *Surg Oncol Clin N Am*. 2016;25:217. [PMID: 26610783]

Fernandez Ranvier GG, Shouhed D, Inabnet WB 3rd. Minimally invasive techniques for resection of pancreatic neuroendocrine tumors. *Surg Oncol Clin N Am*. 2016;25:195. [PMID: 26610782]

Maxwell JE, O'Dorisio TM, Howe JR. Biochemical diagnosis and preoperative imaging of gastroenteropancreatic neuroendocrine tumors. *Surg Oncol Clin N Am*. 2016;25:171. [PMID: 26610781]

Yates CJ, Newey PJ, Thakker RV. Challenges and controversies in management of pancreatic neuroendocrine tumours in patients with MEN1. *Lancet Diabetes Endocrinol*. 2015;3:895. [PMID: 26165399]

Appendix

Normal Hormone Reference Ranges[a,b]

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Adrenocorticotropic hormone (ACTH) stimulation test (cosyntropin test): 0.25 mg of synthetic ACTH$_{1-24}$ (cosyntropin) is administered IV or IM, and serum cortisol is measured at 0, 30, and 60 min. Normal response: peak cortisol >20 µg/dL (>540 nmol/L). A dose of 1 µg of ACTH will give a similar response in the normal individual (see Chapter 9).						
ACTH	Plasma			0.222		Collect in silicone-coated EDTA-containing tubes. Keep iced. Avoid contact with glass during collection and separation. Process immediately. Separate and freeze plasma in plastic tube at −20°C.
		<3 y	Not established		–	
		3-17 y	9-57 ng/L		2-12.7 pmol/L	
		>18 y	6-50 ng/L		1.3-11 pmol/L	
Aldosterone	Serum (fasting)			27.7		Levels in pregnant patients are three to four times higher.
		>18 y				
		Upright 8:00-10:00 AM	<28 ng/dL		<776 pmol/L	
		Upright 4:00-6:00 PM	<21 ng/dL		<582 pmol/L	
		Supine 8:00-10:00 AM	3-16 ng/dL		83.1-443 pmol/L	
		Pediatric				
		1-12 mo	2-70 ng/dL		55.4-1939 pmol/L	
		1-4 y	2-37 ng/dL		55.4-1025 pmol/L	
		5-9 y	<9 ng/dL		<243 pmol/L	
		10-13 y	<21 ng/dL		582 pmol/L	
		14-17 y	<35 mgdL		970 pmol/L	
		Premature Infants (31-35 wk)	<144 ng/dL		3989 pmol/L	
		Term Infants	<217 ng/dL		6011 pmol/L	
	Urine	Sodium diet (100-300 mmol/d)	–		–	Collect 24-h specimen in 10 g boric acid to maintain pH ≤7.5. Refrigerate during collection.
		Children:				
		2-7 y	≤5.7 µg/d			
		8-11 y	≤10.2 µg/d			
		12-16 y	≤15.6 µg/d			
		Adult	≤2.3-21 µg/d			
		Post-fludrocortisone or intravenous saline infusion	≤5.0 µg/d			
Alkaline phosphatase, bone specific	Serum	2-24 mo	25.4-124.0 µg/L	–	–	IRMA; collect following an overnight fast; useful in Paget disease and osteoporosis; assay also known as Ostase; increased levels seen in states of high bone turnover (Paget disease, hyperthyroidism, osteoporosis).
		6-9 y	41-134.6 µg/L			
		10-13 y				
		Males	43.8-177.4 µg/L			
		Females	24.2-154.2 µg/L			
		14-17 y				
		Males	13.7-128 µg/L			
		Females	10.5-75.2 µg/L			
		Adult males				
		18-29 y	8.4-29.3 µg/L			
		30-39 y	7.7-21.3 µg/L			
		40-49 y	7.0-18.3 µg/L			
		60-68 y	7.6-14.9 µg/L			

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
		Adult females				
		18-29 y	4.7-17.8 µg/L			
		30-39 y	5.3-19.5 µg/L			
		40-49 y	5.0-18.8 µg/L			
		50-76 y	5.6-29 µg/L			
Alpha subunit	Serum	Men	≤0.6 ng/mL	–	–	RIA; used to assess gonadotropin hormone subunit secretion in pituitary adenomas.
		Women				
		Premenopausal	≤1.5 ng/mL			
		Postmenopausal	0.9-3.3 ng/mL			
		Hypothyroid	≤3.7 ng/mL			
		Pregnant or hCG-producing tumor	1.8-360 ng/mL			
3α-Androstanediol glucuronide	Serum	Prepubertal children	0.1-0.6 ng/mL	2.14	0.2-1.3 nmol/L	Freeze serum and store at −20°C.
		Tanner Stages II-III				
		Male	0.19-1.64 ng/mL		0.41-3.51 nmol/L	
		Female	0.33-2.44 ng/mL		0.71-5.22 nmol/L	
		Adult				
		Male	2.6-15 ng/mL		5.6-32.1 nmol/L	
		Female	0.6-3.0 ng/mL		1.3-6.4 nmol/L	
Androstenedione	Serum	Men		3.49		
		18-30 y	50-220 ng/dL		175-768 nmol/L	
		31-50 y	40-190 ng/dL		140-663 nmol/L	
		51-60 y	50-220 ng/dL		175-768 nmol/L	
		Women				
		Follicular	35-250 ng/dL		122-873 nmol/L	
		Midcycle	60-285 ng/dL		209-995 nmol/L	
		Luteal	30-235 ng/dL		105-820 nmol/L	
		Postmenopausal	20-75 ng/dL		70-262 nmol/L	
		Children				
		1-12 mo	6-78 ng/dL		21-272 nmol/L	
		1-4 y	5-51 ng/dL		17-178 nmol/L	
		5-9 y	6-115 ng/dL		21-401 nmol/L	
		10-13 y	12-221 ng/dL		42-771 nmol/L	
		14-17 y	22-225 ng/dL		77-785 nmol/L	
		Premature infants (31-35 wk)	17-82 ng/dL		59-286 nmol/L	
		Term Infants	<290 ng/dL		<1012 nmol/L	
		Tanner II-III				
		Male	17-82 ng/dL		59-286 nmol/L	
		Female	43-180 ng/dL		150-628 nmol/L	
		Tanner IV-V				
		Male	57-150 ng/dL		199-523 nmol/L	
		Female	7-68 ng/dL		24-237 nmol/L	
Antidiuretic hormone (ADH; vasopressin)	Plasma	If serum osmolality 283-301 mOsm/kg	1-13 pg/mL	0.925	0.9-12 pmol/L	
Anti-Müllerian hormone	Plasma	Females				
		<14 y	0.3-11.21 ng/mL		–	
		14-19 y	Not established		–	
		20-29 y	0.65-16.40 ng/mL		–	
		30-39 y	0.16-8.43 ng/mL		–	
		40-49 y	<5.20 ng/mL		–	
		>49 y	<2.05 ng/mL		–	
		Males				
		<1 y	101.9-262 ng/mL		–	
		1-6 y	87.3-243.8 ng/mL		–	
		7-11 y	34.3-230.1 ng/mL		–	
		12-17 y	<135.45 ng/mL		–	
		>17 y	1.45-15.27 ng/mL		–	
Angiotensin II	Plasma	Adults	10-50 ng/L	–	–	Collect in EDTA. Centrifuge refrigerated. Store frozen at −70°C.

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
C peptide of insulin	Serum	Fasting	0.8-3.1 ng/mL	0.331	0.26-1.03 nmol/L	Freeze serum at −20°C; stable for 1 d at room temperature and 1 wk at −20°C.
Calcitonin	Serum	Male Female Child ≤6 mo 6 mo-3 y 3-17 y	≤10 pg/mL ≤5 pg/mL ≤41 pg/mL ≤14 pg/mL ≤6 ng/mL	0.293	≤2.93 pmol/L ≤1.46 pmol/L ≤12 pmol/L ≤4.1 pmol/L ≤1.76 pmol/L	Fasting, non-lipemic specimen. Refrigerate, spin down immediately. Store at −20°C.
Calcitonin stimulation test utilizing calcium infusion: 2 mg/kg of calcium in the form of calcium gluconate is administered IV over 1 min. Blood samples for calcitonin are obtained at 1, 2, 5, and 10 min after the infusion. Normal peak values for calcitonin 2 min after calcium infusion: female; <70 ng/L (≤20.5 pmol/L); male: ≤491 ng/L (≤144 pmol/L).							
Calcium, 24-h urinary	Urine		100-300 mg/ 24 h			Collect in 25 mL of 6N HCl. Refrigerate during collection; elevated in primary hyperparathyroidism, Paget disease, multiple myeloma, vitamin D toxicity; reduced in rickets, osteomalacia, familial hypocalciuric hypercalcemia.
Calcium, ionized	Serum	0-1 mo 1-6 mo >6 mo	1-1.5 mmol/L 0.95-1.5 mmol/L 1.16-1.36 mmol/L	–	–	Refrigerated serum must be collected anaerobically in gel barrier tube; elevated in primary hyperparathyroidism, vitamin D toxicity, cancer; reduced in pseudohypoparathyroidism, hypoparathyroidism, and severe vitamin D deficiency.
Calcium, total	Serum	<1 y >1 y	9.0-10.9 mg/dL 8.8-10.3 mg/dL	–	–	Overnight fasting preferred; used to diagnose parathyroid and vitamin D disorders; elevated in primary hyperparathyroidism, vitamin D toxicity; reduced in hypoparathyroidism and vitamin D deficiency.
Catecholamines (fractionated by HPLC)	Plasma	Adults: Norepinephrine Supine Ambulatory Epinephrine Supine Ambulatory Dopamine Supine Ambulatory Total (N+E) Supine Upright	– 112-658 pg/mL 217-1109 pg/mL – ≤50 pg/mL ≤95 pg/mL – ≤10 ng/mL ≤20 ng/mL 123-671 pg/mL 242-1125 pg/mL	0.00591 0.00546 0.00654 0.00591	– 0.66-3.89 nmol/L 1.18-6.55 nmol/L – ≤0.27 nmol/L ≤0.52 nmol/L – ≤0.065 nmol/L ≤0.13 nmol/L 0.73-3.97 nmol/L 1.43-6.64 nmol/L	Collect by intravenous catheter after patient has rested 30 min. Collect and centrifuge under refrigeration; freeze in plastic tube at −20°C.

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
		Children (age 3-15 y):	–	–	–	
		Epinephrine	–	–	–	
		Supine	≤464 pg/mL	0.00546	≤2.54 nmol/L	
		Norepinephrine	–			
		Supine	≤1251 pg/mL	0.00591	≤7.4 nmol/L	
		Dopamine	≤60 pg/mL	0.00654	0.39 nmol/L	
	Urine	Norepinephrine	–	5.91	–	24-h urine preservative: 25 mL of 6N HCl. Freeze aliquot promptly at −20°C.
		3-8 y	5-41 µg/24 h	–	29.5-242.3 nmol/24 h	
		9-12 y	5-50 µg/24 h		29.5-295.8 nmol/24 h	
		13-17 y	12-88 µg/24 h		70.9-526 nmol/24 h	
		>17 y	15-100 µg/24 h		89-591 nmol/24 h	
		Epinephrine	–	5.46		
		3-8 y	1-7 µg/24 h	–	5.46-38.2 nmol/24 h	
		9-12 y	≤8 µg/24 h		≤43.7 nmol/24 h	
		13-17 y	≤11 µg/24 h		≤60 nmol/24 h	
		>17 y	2-24 µg/24 h		11-131 nmol/24 h	
		Total (N + E)		5.91		
		3-8 y	9-51 µg/24 h		53-301 nmol/24 h	
		9-12 y	9-71 µg/24 h		53-420 nmol/24 h	
		13-17 y	13-90 µg/24 h		77-532 nmol/24 h	
		>17 y	26-121 µg/24 h		154-715 nmol/24 h	
		Dopamine		6.54		
		3-8 y	80-378 µg/24 h	–	523-2472 nmol/24 h	
		9-12 y	51-474 µg/24 h		334-3100 nmol/24 h	
		13-17 y	51-645 µg/24 h		334-4218 nmol/24 h	
		>17 y	52-480 µg/24 h		340-3139 nmol/24 h	
Cholecystokinin	Plasma (fasting)		3.9 ng/mL	0.26	1 pmol/L	–
Chorionic gonadotropin, beta subunit (β-hCG)	Serum	Men	≤5 mIU/mL			See Chapter 16 for further details and interpretation.
		Women				
		Premenopausal, Nonpregnant	≤5 mIU/mL			
		Postmenopausal	≤10 mIU/mL			
		Females postconception				
		≤1 wk	5-65 mIU/mL			
		1-2 wk	65-650 mIU/mL			
		2-3 wk	130-6500 mIU/mL			
		3-4 wk	650-13,000 mIU/mL			
		4-5 wk	1300-65,000 mIU/mL			
		5-6 wk	13,000-130,000 mIU/mL			
		6-8 wk	19,500-260,000 mIU/mL			
		2-3 mo	13,000-130,000 mIU/mL			
Chromogranin A	Serum	Adults	1.9-15 ng/mL			–
Collagen type 1 C-telopeptide (CTx)	Serum	Adults	–			–
		Men				Frozen serum, collected between 8 AM and 10 AM. Used to evaluate osteoporosis and Paget disease.
		18-29 y	87-1200 pg/mL			
	–	30-39 y	70-780 pg/mL			
		40-49 y	60-700 pg/mL			
		50-68 y	87-345 pg/mL			
		Women				
		18-29 y	64-640 pg/mL			
		30-39 y	60-650 pg/mL			
		40-49 y	40-465 pg/mL			

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Collagen cross-linked N-telopeptide (NTx), urine	2 mL refrigerated urine	Men		–	–	Monitor response to treatment in patients with osteoporosis; diagnose high bone turnover state; possibly elevated in Paget disease of bone, hyperthyroidism, osteoporosis.
		Second AM voided urine				
		18-29 y	12-99 nmol BCE/mmol creatinine			
		30-59 y	9-60 nmol BCE/mmol creatinine			
		24-h urine				
		18-29 y	5-88 nmol BCE/mmol creatinine			
		30-39 y	7-51 nmol BCE/mmol creatinine			
		40-49 y	5-47 nmol BCE/mmol creatinine			
		50-60 y	6-43 nmol BCE/mmol creatinine			
		Women				
		Second AM voided urine Premenopausal	4-64 nmol BCE/mmol creatinine			
		24-h urine >18 y old Premenopausal	5-79 nmol BCE/mmol creatinine			

Corticotropin-releasing hormone (CRH) test: Ovine CRH in a dose of 1 μg/kg is administered IV. Blood samples for ACTH and cortisol determinations are taken at 15, 30, and 60 min. The peak ACTH response of >10 pg/mL (>2.2 pmol/L) occurs at 15 min. The peak cortisol response of >10 μg/dL (>280 nmol/L) occurs at 30-60 min (see Chapter 4).

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Corticotropin-releasing hormone	Plasma	Men and nonpregnant women	≤34 pg/mL	2.0	≤68 pmol/L	Markedly elevated at term pregnancy.
		Pregnant women				
		1st trimester	≤40 pg/mL		≤80 pmol/L	
		2nd trimester	≤153 pg/mL		≤306 pmol/L	
		3rd trimester	≤847 pg/mL		≤1694 pmol/L	
		Cord blood	≤338 pg/mL		≤676 pmol/L	
Cortisol	Serum	Adult		27.59		Collect and process under refrigeration. Spin down immediately. Salivary cortisol is in equilibrium with free cortisol and may be used as an index to free cortisol. Reference range may vary with method and laboratory.
		Total				
		7-9 AM	4-22 μg/dL		193-607 nmol/L	
		3-5 PM	3-17 μg/dL		83-469 nmol/L	
		Peak post-ACTH	>20 μg/dL		551.8 nmol/L	
		Children				
		Premature (31-35 wk)	≤15 μg/dL		≤414 nmol/L	
		Term Infants (3 d)	≤14 μg/dL		≤386 nmol/L	
		1-17 y	2-17 μg/dL		55-469 nmol/L	
		Free				
		Adult				
		8-10 AM	0.07-0.93 μg/dL		2-25.6 nmol/L	
		4-6 PM	0.04-0.45 μg/dL		1.1-12.4 nmol/L	
		10-11 PM	0.04-0.35 μg/dL		1.1-9.7 nmol/L	
	Saliva	Adult				
		8-10 AM	0.04-0.56 μg/dL		1.1-15.6 nmol/L	
		4-6 PM	<0.15 μg/dL		<4.2 nmol/L	
		10-11 PM	<0.09 μg/dL		<2.5 nmol/L	

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Dehydroepi-androsterone (DHEA)	Urine (free)	24-h specimen RIA Adults Children	4-50 µg/24 h	2.76	11-138 nmol/24 h	Collect 24-h speci-men with 8 g of boric acid or 10 mL of 6N HCl as preservative. Liquid chromatography/ tandem mass spectrometry.
		1-4 y	0.9-8.2 µg/24 h		2.5-22.6 nmol/24 h	
		5-9 y	1-30 µg/24 h		2.8-82.8 nmol/24 h	
		10-13 y	1-45 µg/24 h		2.8-124.2 nmol/24 h	
		14-17 y	3-55 µg/24 h		8.3-151.8 nmol/24 h	
	Serum (fasting preferred)	1-5 y	<68 ng/dL	0.347	<23.5 nmol/L	Separate serum immediately and store at −20°C.
		6-7 y	<111 ng/dL		<38.5 nmol/L	
		8-10 y	<186 ng/dL		<64.5 nmol/L	
		11-12 y	<202 ng/dL		<70 nmol/L	
		13-14 y	<319 ng/dL		<111 nmol/L	
		15-16 y	39-481 ng/dL		13.5-166.9 nmol/L	
		17-19 y	40-491 ng/dL		13.9-170.3 nmol/L	
		20-50 y	31-701 ng/dL		10.8-243.2 nmol/L	
Dehydroepi-androsterone sulfate (DHEAS)	Serum (fasting preferred)	Male		0.0272		Stable 72 h at 40°C. Store at −20°C.
		18-29 y	110-510 µg/dL		3.0-13.9 µmol/L	
		30-39 y	110-370 µg/dL		3.0-10.1 µmol/L	
		40-49 y	45-345 µg/dL		1.2-9.4 µmol/L	
		50-59 y	25-240 µg/dL		0.7-6.5 µmol/L	
		60-69 y	25-95 µg/dL		0.7-2.6 µmol/L	
		70-90 y	≤76 µg/dL		≤2.1 µmol/L	
		Female				
		18-29 y	45-320 µg/dL		1.2-8.7 µmol/L	
		30-39 y	40-325 µg/dL		1.1-8.8 µmol/L	
		40-49 y	25-220 µg/dL		0.7-6.0 µmol/L	
		50-59 y	15-170 µg/dL		0.4-4.6 µmol/L	
		60-69 y	≤186 µg/dL		≤5.0 µmol/L	
		70-90 y	≤91 µg/dL		≤2.5 µmol/L	
		Pediatrics Premature Infants				
		26-28 wk, day 4	123-882 µg/dL		3.3-24 µmol/L	
		31-35 wk, day 4	122-710 µg/dL		3.3-19 µmol/L	
		Full Term Infants				
		26-26 wk, day 4	123-882 µg/dL		3.3-24 µmol/L	
		31-35 wk, day 4	122-710 µg/dL		3.3-19 µmol/L	
		Prepubertal				
		1-5 y	<5-5 µg/dL		<0.1-0.1 µmol/L	
		6-7 y	9-72 µg/dL		0.2-2.0 µmol/L	
		8-10 y	13-115 µg/dL		0.35-3.1 µmol/L	
		Tanner Stage Males				
		I	13-83 µg/dL		0.35-2.3 µmol/L	
		II	42-109 µg/dL		1.1-3.0 µmol/L	
		III	48-200 µg/dL		1.3-5.4 µmol/L	
		IV	102-385 µg/dL		2.8-10.5 µmol/L	
		V	120-370 µg/dL		3.2-10.1 µmol/L	
		Females				
		I	99-144 µg/dL		2.7-3.9 µmol/L	
		II	32-129 µg/dL		0.9-3.5 µmol/L	
		III	32-226 µg/dL		0.9-6.1 µmol/L	
		IV	58-260 µg/dL		1.6-7.1 µmol/L	
		V	44-248 µg/dL		1.2-6.7 µmol/L	
Deoxycorticos-terone (DOC)	Serum (fasting preferred)			30.26		Process immediately. Store at −20°C.
		Cord blood	111-372 ng/dL		3359-11257 pmol/L	
		1 wk-12 mo	7-49 ng/dL		212-1483 pmol/L	
		Prepubertal (1-10 y)	2-34 ng/dL		60.5-1029 pmol/L	
		<18 y (8 AM)	2-19 ng/dL		60.5-574.9 pmol/L	

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
11-Deoxycortisol	Serum			0.02887		Process immediately. Store at −20°C. Early morning specimen preferred.
		Cord blood	295-554 ng/dL		8.5-16 nmol/L	
		Premature Infants	48-579 ng/dL		1.4-16.7 nmol/L	
		Full Term Infants	13-147 ng/dL		0.4-4.2 nmol/L	
		1-12 mo	<156 ng/dL		<4.5 nmol/L	
		Prepubertal (1-10 y)	20-155 ng/dL		0.6-4.8 nmol/L	
		>18 y	12-158 ng/dL		0.3-4.6 nmol/L	

Dexamethasone suppression test (low dose) for the diagnosis of Cushing syndrome (see Chapter 9): Obtain a baseline serum cortisol at 0700-0800 h. Administer 1 mg dexamethasone orally at 2300 h that evening and obtain another serum cortisol at 0700-0800 h the following morning. **Interpretation:** A normal response (normal suppressibility) is a reduction of the post-dexamethasone serum cortisol to ≤1.8 μg/dL (≤50 nmol/L).

Dexamethasone suppression test (high dose) for the differential diagnosis of Cushing syndrome (see Chapter 9): Obtain a baseline serum cortisol at 0700-0800 h. Administer 8 mg dexamethasone orally at 2300 h that evening and obtain another serum cortisol at 0700-0800 h the following morning. **Interpretation:** A reduction of the post-dexamethasone serum cortisol to ≤50% of the baseline cortisol indicates suppressibility.

Dexamethasone-CRH test: Administer dexamethasone, 0.5 mg every 6 h orally for eight doses, followed by CRH, 1 μg/kg IV 2 h after the last dose of dexamethasone. Plasma cortisol is obtained 15 min after CRH. Normal: ≤1.4 μg/dL (≤38.6 nmol/L). (See Chapter 9.)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Dihydrotesterone (male)	Serum			0.0344		Separate serum within 1 h after collection and store at −20°C.
		Cord blood	≤2-8 ng/dL		<.06-.3 nmol/L	
		Premature Infants	10-53 ng/dL		0.3-1.8 nmol/L	
		Full Term Infants	5-60 ng/dL		0.2-2.1 nmol/L	
		30-60 d	12-85 ng/dL		≤0.41-2.9 nmol/L	
		Prepubertal	≤3 ng/dL		<0.1 nmol/L	
		Tanner II-III	3-33 ng/dL		0.1-1.1 nmol/L	
		Tanner IV-V	22-65 ng/dL		0.8-2.2 nmol/L	
		Adult	30-85 ng/dL		1-3.0 nmol/L	
Dihydrotesterone (female)	Serum	Cord blood	≤2-8 ng/dL	0.0344	≤0.07-0.27 nmol/L	
		Premature Infants	2-13 ng/dL		0.07-0.45 nmol/L	
		Full Term Infants	≤2-15 ng/dL		≤0.07-0.52 nmol/L	
		30-60 d	<3 ng/dL		<0.1 nmol/L	
		Prepubertal	<3 ng/dL		<0.1 nmol/L	
		Tanner II-III	5-19 ng/dL		0.2-0.7 nmol/L	
		Tanner IV-V	4-18 ng/dL		0.14-0.6 nmol/L	
		Adult	4-22 ng/dL		0.14-7.6 nmol/L	
Erythropoietin	Serum	Adult	4.1-19.5 mU/mL	1.00	4.1-19.5 mU/mL	
		Child				
		3 wk-2 mo	5.0-13.0 mU/mL		5.0-13.0 mU/mL	
		3 mo-16 y	9.0-28.0 mU/mL		9.0-28.0 mU/mL	
Estradiol, Ultrasensitive	Serum	Male		3.67		
		1-9 y	≤4 pg/mL		≤15 pmol/L	
		10-11 y	≤12 pg/mL		≤44 pmol/L	
		12-14 y	≤24 pg/mL		≤88 pmol/L	
		15-17 y	≤31 pg/mL		≤114 pmol/L	
		>17 y	≤29 pg/mL		≤106 pmol/L	
		Female				
		1-9 y	≤16 pg/mL		≤59 pmol/L	
		10-11 y	≤65 pg/mL		≤239 pmol/L	
		12-14 y	≤142 pg/mL		≤521 pmol/L	
		15-17 y	≤283 pg/mL		≤1039 pmol/L	
		Follicular	39-375 pg/mL		143-1376 pmol/L	
		Midcycle peak	96-762 pg/mL		352-2797 pmol/L	
		Luteal	48-440 pg/mL		176-1614 pmol/L	
		Postmenopausal	≤10 pg/mL		≤37 pmol/L	
Estradiol, free		Adult male	≤0.45 pg/mL	3.67	≤1.7 pmol/mL	
		Adult female				
		Follicular	0.43-5.03 pg/mL		1.6-18.5 pmol/mL	
		Midcycle	0.72-5.89 pg/mL		2.6-21.6 pmol/mL	
		Luteal	0.40-5.55 pg/mL		1.5-20.4 pmol/mL	
		Postmenopausal	≤0.38 pg/mL		≤1.4 pmol/mL	

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Estrone	Serum	Adult male	≤68 ng/L	3.70	≤251 pmol/L	
		Postpubertal female				
		Follicular phase	10-138 ng/L		37-511 pmol/L	
		Midcycle	49-268 ng/L		181-992 pmol/L	
		Luteal phase	16-173 ng/L		59-640 pmol/L	
		Postmenopausal	≤65 ng/L		≤240 pmol/L	
		Pediatric males				
		1-9 y	<10 pg/mL		<37 pmol/L	
		10-11 y	<12 pg/mL		<44 pmol/L	
		12-14 y	<28 pg/mL		<104 pmol/L	
		15-17 y	<64 pg/mL		<237 pmol/L	
		Pediatric females				
		1-9 y	<32 pg/mL		<118 pmol/L	
		10-11 y	<72 pg/mL		<266 pmol/L	
		12-14 y	<75 pg/mL		<278 pmol/L	
		15-17 y	<188 pg/mL		<696 pmol/L	
Follicle stimulating hormone	Serum or plasma (heparin)	Adult males (>18 y)	0.6-14.7 IU/L	1.00	0.6-14.7 IU/L	
		Adult females (>18 y)				
		Follicular phase	2.5-10.2 IU/L		2.5-10.2 IU/L	
		Midcycle peak	3.4-33.4 IU/L		3.4-33.4 IU/L	
		Luteal phase	1.5-9.1 IU/L		1.5-9.1 IU/L	
		Pregnant	<0.3 IU/L		<0.3 IU/L	
		Postmenopausal	23-116 IU/L		23-116 IU/L	
		Pediatric males				
		0-9 y	≤3.0 IU/L		≤3.0 IU/L	
		10-13 y	0.3-4.0 IU/L		0.3-4.0 IU/L	
		14-17 y	0.4-7.4 IU/L		0.4-7.4 IU/L	
		Pediatric females				
		0-8 y	0.5-4.5 IU/L		0.5-4.5 IU/L	
		9-13 y	0.4-6.5 IU/L		0.4-6.5 IU/L	
		14-17 y	0.8-8.5 IU/L		0.8-8.5 IU/L	
Gastrin	Serum	Adults	≤100 pg/mL	0.475	≤48 pmol/L	Overnight fast required. Store at −20°C.
		Children 5-7 y	13-64 pg/mL		6.2-30 pmol/L	
Glucagon	Plasma	Adults	≤60 pg/mL	0.287	≤17.2 pmol/L	Centrifuge immediately under refrigeration. Store in plastic vial with 0.5 mL aprotinin (10,000 KIU/mL) at −20°C. Overnight fast required.
		Children				
		Cord blood	≤215 pg/mL		≤62 pmol/L	
		Day 1	≤240 pg/mL		≤69 pmol/L	
		Day 2	≤400 pg/mL		≤115 pmol/L	
		Day 3	≤420 pg/mL		≤121 pmol/L	
		Day 4-14	≤148 pg/mL		≤42 pmol/L	
Growth hormone	Serum	Adult		46.5		Store at −20°C. **Note:** GH values fluctuate widely, and functional tests must be utilized for diagnosis of GH deficiency or excess. See Chapter 4 for details of suppression and stimulation tests for GH excess or deficiency.
		Males	<3.1 ng/mL		<144 pmol/L	
		Females	<8.1 ng/mL		<377 pmol/L	
		Post-75-g glucose load	≤1 ng/mL		≤46 pmol/L	
		Post overnight fast	0-6 ng/mL		0-279 pmol/L	
		Children				
		1-2 d	5-53 ng/mL		232-2465 pmol/L	
		2-7 d	5-27 ng/mL		232-1256 pmol/L	
		31 d-11 mo	2-10 ng/mL		93-465 pmol/L	
Growth hormone–binding protein	Serum	Adults	400-4260 pmol/L	1.0	400-4260 pmol/L	Store at −20°C.
		Children				
		3-8 y	320-3820 pmol/L		320-3820 pmol/L	
		9-13 y	240-2890 pmol/L		240-2890 pmol/L	
		14-17 y	290-3140 pmol/L		290-3140 pmol/L	

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Growth hormone–releasing hormone	Plasma	Adults Children (4-14 y)	≤41 pg/mL 6.8-19 pg/mL	1.0	≤41 pg/mL 6.8-19 pg/mL	Store at –20°C.
Hemoglobin A₁c	Whole blood	Nondiabetic	4.3-5.6%			Collection in EDTA. Keep refrigerated.
Homovanillic acid	Urine	3-8 y 9-12 y 13-17 y >17 y	0.5-6.7 mg/24 h 1.1-6.8 mg/24 h 1.4-7.2 mg/24 h 1.6-7.5 mg/24 h	5.49	2.7-36.8 µmol/24 h 6.0-37.3 µmol/24 h 7.7-39.5 µmol/24 h 8.8-41.2 µmol/24 h	Preservative: 10 mL of 6N HCl. Avoid alcohol, coffee, tea, tobacco, and strenuous exercise for 2 wk pretest.
17-Hydroxy-corticoids	Urine	Males, age 1-20 y Adult males Females, age 1-20 y Adult females	0.5-10 mg/24 h 3-10 mg/24 h 0.5-7 mg/24 h 2-6 mg/24 h	2.76	1.4-28 µmol/24 h 8.3-28 µmol/24 h 1.4-19 µmol/24 h 5.5-16.6 µmol/24 h	Preservative: 10 mL of 6N HCl.
5-Hydroxy-indoleacetic	Urine	Age 2-10 y Age >10 y	≤8 mg/24 h ≤6 mg/24 h	5.23	≤41.8 µmol/24 h ≤31.4 µmol/24 h	Preservative: 10 mL of 6N HCl. Refrigerate during 24-h collection. For 48 h prior to and during collection, avoid avocados, and alcohol.

18-Hydroxy-corticosterone	Serum	Adult Supine (8-10 AM)
		Ambulatory (8-10 AM)
		Children
		  Premature infants (31-35 wk)
		  Term infants (3 d)
			4-37 ng/dL
			5-80 ng/dL
			
			≤380 ng/dL
			
			≤942 ng/dL
				27.51	110-1018 pmol/L
					138-2201 pmol/L
					
					≤10,454 pmol/L
					
					≤25,914 pmol/L
						Refrigerate.

17-Hydroxy-pregnenolone	Serum	Adults
		  Post-ACTH
		Children
		  Premature infants
		  Term infants
		  1-12 mo
		  1-10 y
		  Pubertal
			53-357 ng/dL
			290-910 ng/dL
			
			64-2380 ng/dL
			10-829 ng/dL
			36-540 ng/dL
			15-221 ng/dL
			44-235 ng/dL
				0.0301	1.6-10.7 nmol/L
					8.7-27.4 nmol/L
					
					1.9-71.6 nmol/L
					0.3-25 nmol/L
					1.1-16.3 nmol/L
					0.5-6.7 nmol/L
					1.3-7.1 nmol/L
						Process immediately. Store at –20°C.

17-Hydroxy-progesterone	Serum (fasting preferred)	Males
		  18-30 y
		  31-40 y
		  41-50 y
		  51-60 y
		Females
		  Follicular phase
		  Midcycle phase
		  Luteal phase
		  Postmenopausal
		Children
		  Premature infants
		  Term infants
		  1-12 mo
		  1-4 y
		  5-9 y
		  10-13 y
		  14-17 y
		Post-ACTH
		  1-12 mo
		  1-5 y
		  6-12 y
			
			32-307 ng/dL
			42-196 ng/dL
			33-195 ng/dL
			37-129 ng/dL
			
			≤185 ng/dL
			≤225 ng/dL
			≤285 ng/dL
			≤45 ng/dL
			
			≤360 ng/dL
			≤420 ng/dL
			11-170 ng/dL
			4-115 ng/dL
			≤90 ng/dL
			≤169 ng/dL
			16-283 ng/dL
			
			85-465 ng/dL
			50-350 ng/dL
			75-220 ng/dL
				0.0303	
					1-9.3 nmol/L
					1.3-5.9 nmol/L
					1-5.9 nmol/L
					1.1-3.9 nmol/L
					
					≤5.6 nmol/L
					≤6.8 nmol/L
					≤8.6 nmol/L
					≤1.4 nmol/L
					
					≤10.9 nmol/L
					≤12.7 nmol/L
					0.3-5.2 nmol/L
					0.1-3.5 nmol/L
					≤2.7 nmol/L
					≤5.1 nmol/L
					0.5-8.6 nmol/L
					
					2.6-14.1 nmol/L
					1.5-10.6 nmol/L
					2.3-6.7 nmol/L	

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Insulin, total (free and bound)	Serum	Fasting	3.3-22.1 µU/mL (0.15-0.884 ng/mL)	172.1 (ng/mL → pmol/L)	25.8-152.1 pmol/L	Cold centrifuge. Freeze at –20°C. Must always be evaluated in light of simultaneous plasma glucose measurement.
Insulin with oral glucose tolerance test	Serum	1 h 2 h	50-130 µU/mL (2-5.2 ng/mL) ≤30 µU/mL (≤1.2 ng/mL)	172.1 (ng/mL → pmol/L)	344-895 pmol/L ≤207 pmol/L	Cold centrifuge. Freeze at –20°C.
Insulin autoantibody	Serum		<0.4 U/mL			
Insulin-like growth factor-I (IGF-I)	Serum			1.0		Store refrigerated. Elevated in acromegaly and gigantism; reduced in isolated GH deficiency. Levels in the ultrasensitive IGF-I assay are shown first followed by routinely performed IGF-I assays in adults over 18 y.

IGF-I detail:

Ages, Conditions, etc	Conventional Units (Term)	(Preterm)	SI Units (Term)	(Preterm)
Newborns and Infants				
Birth	15-109 ng/mL	21-93	15-109 µg/L	21-93
1 day-2 mo	15-109 ng/mL	23-163	15-109 µg/L	23-163
3-4 mo	7-124 ng/mL	23-171	7-124 µg/L	23-171
5-6 mo	7-93 ng/mL	15-132	7-93 µg/L	15-132
7-12 mo	15-101 ng/mL	15-179	15-101 µg/L	15-179
Males				
1-2 y	30-122 ng/mL		30-122 µg/L	
Males – Tanner I				
3 y	20-141 ng/mL		20-141 µg/L	
4 y	25-157 ng/mL		25-157 µg/L	
5 y	30-174 ng/mL		30-174 µg/L	
6 y	37-192 ng/mL		37-192 µg/L	
7 y	44-211 ng/mL		44-211 µg/L	
8 y	52-231 ng/mL		52-231 µg/L	
9 y	61-252 ng/mL		61-252 µg/L	
10 y	71-275 ng/mL		71-275 µg/L	
11 y	82-299 ng/mL		82-299 µg/L	
12 y	93-324 ng/mL		93-324 µg/L	
13 y	106-350 ng/mL		106-350 µg/L	
14 y	120-377 ng/mL		120-377 µg/L	
15 y	127-391 ng/mL		127-391 µg/L	
Males – Tanner II & III				
8 y	39-264 ng/mL		39-264 µg/L	
9 y	52-304 ng/mL		52-304 µg/L	
10 y	67-347 ng/mL		67-347 µg/L	
11 y	86-393 ng/mL		86-393 µg/L	
12 y	106-443 ng/mL		106-443 µg/L	
13 y	130-497 ng/mL		130-497 µg/L	
14 y	156-554 ng/mL		156-554 µg/L	
15 y	185-616 ng/mL		185-616 µg/L	
16 y	201-648 ng/mL		201-648 µg/L	
Males – Tanner IV & V				
11 y	277-673 ng/mL		277-673 µg/L	
12 y	365-652 ng/mL		365-652 µg/L	
13 y	241-612 ng/mL		241-612 µg/L	
14 y	220-574 ng/mL		220-574 µg/L	
15 y	199-537 ng/mL		199-537 µg/L	
16 y	180-501 ng/mL		180-501 µg/L	
17 y	161-467 ng/mL		161-467 µg/L	
18 y	144-434 ng/mL		144-434 µg/L	

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
		Males				
		19-20 y	281-510 ng/mL		281-510 µg/L	
		21-30 y	155-432 ng/mL		155-432 µg/L	
		31-40 y	132-333 ng/mL		132-333 µg/L	
		41-50 y	121-237 ng/mL		121-237 µg/L	
		51-60 y	68-245 ng/mL		68-245 µg/L	
		61-70 y	60-220 ng/mL		60-220 µg/L	
		71-80 y	36-215 ng/mL		36-215 µg/L	
		Females				
		1-2 y	56-144 ng/mL		56-144 µg/L	
		Females – Tanner I				
		3 y	26-162 ng/mL		26-162 µg/L	
		4 y	32-179 ng/mL		32-179 µg/L	
		5 y	39-198 ng/mL		39-198 µg/L	
		6 y	47-217 ng/mL		47-217 µg/L	
		7 y	55-238 ng/mL		55-238 µg/L	
		8 y	64-259 ng/mL		64-259 µg/L	
		9 y	74-282 ng/mL		74-282 µg/L	
		10 y	85-306 ng/mL		85-306 µg/L	
		11 y	97-332 ng/mL		97-332 µg/L	
		12 y	110-358 ng/mL		110-358 µg/L	
		Females – Tanner II				
		8 y	89-369 ng/mL		89-369 µg/L	
		9 y	96-399 ng/mL		96-399 µg/L	
		10 y	104-431 ng/mL		104-431 µg/L	
		11 y	112-466 ng/mL		112-466 µg/L	
		12 y	121-504 ng/mL		121-504 µg/L	
		13 y	131-545 ng/mL		131-545 µg/L	
		14 y	136-566 ng/mL		136-566 µg/L	
		Females – Tanner III				
		9 y	192-568 ng/mL		192-568 µg/L	
		10 y	192-568 ng/mL		192-568 µg/L	
		11 y	192-568 ng/mL		192-568 µg/L	
		12 y	192-568 ng/mL		192-568 µg/L	
		13 y	192-568 ng/mL		192-568 µg/L	
		14 y	192-568 ng/mL		192-568 µg/L	
		15 y	192-568 ng/mL		192-568 µg/L	
		Females – Tanner IV & V				
		10 y	279-664 ng/mL		279-664 µg/L	
		11 y	268-646 ng/mL		268-646 µg/L	
		12 y	248-612 ng/mL		248-612 µg/L	
		13 y	229-579 ng/mL		229-579 µg/L	
		14 y	211-547 ng/mL		211-547 µg/L	
		15 y	194-516 ng/mL		194-516 µg/L	
		16 y	177-487 ng/mL		177-487 µg/L	
		17 y	162-458 ng/mL		162-458 µg/L	
		18 y	147-430 ng/mL		147-430 µg/L	
		Females				
		19-20 y	217-475 ng/mL		217-475 µg/L	
		21-30 y	87-368 ng/mL		87-368 µg/L	
		31-40 y	106-368 ng/mL		106-368 µg/L	
		41-50 y	118-298 ng/mL		118-298 µg/L	
		51-60 y	53-287 ng/mL		53-287 µg/L	
		61-70 y	75-263 ng/mL		75-263 µg/L	
		71-80 y	54-205 ng/mL		54-205 µg/L	
		Reference range for standard sensitivity assay.				

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
		Either males or females				
		18-19.9 y	108-548 ng/mL		108-548 µg/L	
		20-24.9 y	83-456 ng/mL		83-456 µg/L	
		25-29.9 y	63-373 ng/mL		63-373 µg/L	
		30-39.9 y	53-331 ng/mL		53-331 µg/L	
		40-49.9 y	52-328 ng/mL		52-328 µg/L	
		50-59.9 y	50-317 ng/mL		50-317 µg/L	
		60-69.9 y	41-279 ng/mL		41-279 µg/L	
		70-79.9 y	34-245 ng/mL		34-245 µg/L	
		>80 y	34-246 ng/mL		34-246 µg/L	
Insulin-like growth factor-II (IGF-II)	Serum	Pre-puberty Puberty >18 y old	334-642 µg/L 245-737 µg/L 288-736 µg/L	1.0		Store serum at −20 C. Elevated in mesenchymal tumors with hypoglycemia; overnight fasting preferred.
Insulin-like growth factor–binding protein-III (IGFBP III)	Serum	Premature	Range	1.0	Mean	Separate serum within 1 h. Free serum in plastic vial at −20°C. Major carrier of IGFs, transporting ~95% of IGF-I and IGF-II. IGF-BP3 is GH-responsive and reflects GH actions: elevated in acromegaly and low in hypopituitarism and GH deficiency. Can be used to discriminate among causes of short stature.
		0-1 mo	0.3-1.4 mg/L		0.9 mg/L	
		2-3 mo	0.9-2.3 mg/L		1.6 mg/L	
		4-5 mo	0.4-2.2 mg/L		1.5 mg/L	
		6-11 mo	1.0-2.3 mg/L		1.5 mg/L	
		Full Term				
		0-1 mo	0.4-1.7 mg/L		0.9 mg/L	
		2-3 mo	0.5-2.1 mg/L		1.3 mg/L	
		4-5 mo	0.6-2.4 mg/L		1.4 mg/L	
		6-11 mo	0.5-2.4 mg/L		1.4 mg/L	
		1-4 y	0.8-3.0 mg/L		2.1 mg/L	
		5-6 y	1.5-3.4 mg/L		2.4 mg/L	
		7-8 y	2.1-4.2 mg/L		3.0 mg/L	
		9-11 y	2.0-4.8 mg/L		3.3 mg/L	
		12-13 y	2.1-6.2 mg/L		3.8 mg/L	
		14-15 y	2.2-5.9 mg/L		4.2 mg/L	
		16-18 y	2.5-4.8 mg/L		3.5 mg/L	
		19-30 y	2.0-4.2 mg/L		3.0 mg/L	
		31-70 y	1.9-3.6 mg/L		2.7 mg/L	
Islet cell antibody (ICA)	Serum	Adults and children	Negative (≤1.25 JDF units)	1.0	Negative (≤1.25 JDF units)	Room temperature (2 d) or refrigerated serum sample (2 wk). ICA includes antibodies directed against GAD (glutamic acid decarboxylase), IA-2, insulin, GM2-1, and other cell surface proteins. ICA is present during the prediabetic phase and predicts the development of type 1 diabetes. Titers are compared to a single international reference standard reported in JDF (Juvenile Diabetes Foundation) units.

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
17-Ketosteroids	Urine	Males			3.47	Collect urine for
		0	<1.0 mg/d		<3. 47 µmol/d	24 h with
		1 y	<2.0 mg/d		<7.94 µmol/d	10 g of boric acid,
		5 y	<3.0 mg/d		<10.41 µmol/d	30 mL of 6N HCl
		9 y	3.0-10.0 mg/d		10.41-34.7 µmol/d	or 25 mL 50% ace-
		13 y	5.0-12.0 mg/d		17.35-41.64 µmol/d	tic acid and refrig-
		17 y	9.0-22.0 mg/d		31.23-76.34 µmol/d	erate during
		21 y	8.0-20.0 mg/d		27.76-69.4 µmol/d	collection. Submit
		>21 y	8.0-20.0 mg/d		27.76-69.4 µmol/d	20 mL sample.
		Females				
		0	<1.0 mg/d		<3. 47 µmol/d	
		1 y	<2.0 mg/d		<7.94 µmol/d	
		5 y	<3.0 mg/d		<10.41 µmol/d	
		9 y	3.0-10.0 mg/d		10.41-34.7 µmol/d	
		13 y	5.0-12.0 mg/d		17.35-41.64 µmol/d	
		17 y	6.0-15.0 mg/d		20.82-180.6 µmol/d	
		21 y	(not available)		(not available)	
		>21 y	6.0-15.0 mg/d		20.82-180.6 µmol/d	
Leptin	Serum	Adult males		1.0		Refrigerated serum
		(BMI 18-25, age 18-61 y)	1.2-9.5 ng/mL		1.2-9.5 µg/L	sample required
		(BMI 25-30, age 19-60 y)	1.6-20.9 ng/mL		1.6-20.9 µg/L	for RIA.
		Adult females				
		(BMI 18-25, age 18-61 y)	4.1-25.0 ng/mL		4.1-25.0 µg/L	
		(BMI 25-30, age 19-60 y)	13.1-40.8 ng/mL		13.1-40.8 µg/L	
		Children				
		Prepubertal males	1.6-10.8 ng/mL		1.6-10.8 µg/L	
		Tanner II-III males	2.1-11.6 ng/mL		2.1-11.6 µg/L	
		Tanner IV-V males	3.4-10.2 ng/mL		3.4-10.2 µg/L	
		Prepubertal females	1.7-10.6 ng/mL		1.7-10.6 µg/L	
		Tanner II-III females	2.6-11.5 ng/mL		2.6-11.5 µg/L	
		Tanner IV-V females	3.4-13.0 ng/mL		3.4-13.0 µg/L	
Luteinizing hormone	Plasma or serum	Males		1.0		Test measures the sum of LH and hCG; high hCG levels in preg-
		Tanner Stage/Age (y)	Range			nancy or tropho-
		I <9.8	0.02-0.3 mIU/mL		0.02-0.3 IU/L	blastic disease
		II 9.8-14.5	0.2-4.9 mIU/mL		0.2-4.9 IU/L	cross-react in the
		III 10.7-15.4	10.7-15.4 mIU/mL		10.7-15.4 IU/L	assay, giving
		IV 11.8-17.3	11.8-17.3 mIU/mL		11.8-17.3 IU/L	falsely high LH
		Adults				levels. Third-
		20-70 y	0.7-7.9 mIU/mL		0.7-7.9 IU/L	generation (ultra-
		>70 y	3.1-34.6 mIU/mL		3.1-34.6 IU/L	sensitive) LH assay
		Females				is sensitive to
		Tanner Stage/Age (y)				0.03 mIU/mL and
		I <9.8	0.02-0.18 mIU/mL		0.02-0.18 IU/L	is more appropri-
		II 9.2-13.7	0.02-4.7 mIU/mL		0.02-4.7 IU/L	ate for use in
		III 10.0-14.4	0.1-12.0 mIU/mL		0.1-12.0 IU/L	children. Freeze
		IV 10.7-18.6	0.4-11.7 mIU/mL		0.4-11.7 IU/L	specimen at −20°C.

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
		Adults				
		Follicular phase	1.9-12.5 mIU/mL		1.9-12.5 IU/L	
		Mid-cycle peak	8.7-76.3 mIU/mL		8.7-76.3 IU/L	
		Luteal phase	0.5-16.9 mIU/mL		0.5-16.9 IU/L	
		Pregnant	<0.1-1.5 mIU/mL		<0.1-1.5 IU/L	
		Postmenopausal	15.9-54.0 mIU/mL		15.9-54.0 IU/L	
Metanephrines, fractionated, urine	Urine	Metanephrines		5.07		5 mL aliquot of urine; abstain from medication, tobacco, tea, and coffee for 3 d prior to collection. Common antihypertensive medications cause no interference (diuretics, ACE-inhibitors, calcium-channel blockers, alpha and beta blockers). Medications which are alpha agonists (Aldomet) or alpha blockers (dibenzyline) should be avoided for 18-24 h before specimen collection.
		3 mo-4 y	25-117 µg/24 h		126.8-593.2 nmol/24 h	
		5-9 y	11-139 µg/24 h		55.8-704.7 nmol/24 h	
		10-13 y	51-275 µg/24 h		258.6-1394.3 nmol/24 h	
		14-17 y	40-189 µg/24 h		202.8-958.2 nmol/24 h	
		18-29 y	25-222 µg/24 h		126.8-1125.6 nmol/24 h	
		30-39 y	36-190 µg/24 h		182.5-963.3 nmol/24 h	
		40-49 y	58-203 µg/24 h		294.1-1029.2 nmol/24 h	
		≥50 y	90-315 µg/24 h		456.3-1597.1 nmol/24 h	
		Normetanephrines				
		3 mo-4 y	54-249 µg/24 h		273.8-1262.4 nmol/24 h	
		5-9 y	31-398 µg/24 h		157.2-2018 nmol/24 h	
		10-13 y	67-503 µg/24 h		339.7-2550 nmol/24 h	
		14-17 y	69-531 µg/24 h		349.8-2692 nmol/24 h	
		18-29 y	40-412 µg/24 h		202.8-2089 nmol/24 h	
		30-39 y	35-482 µg/24 h		177.4-2444 nmol/24 h	
		40-49 y	88-649 µg/24 h		446.2-3290 nmol/24 h	
		≥50 y	122-676 µg/24 h		618.5-3427 nmol/24 h	
		Metanephrines, total				
		3 mo-4 y	79-345 µg/24 h		400.5-1749 nmol/24 h	
		5-9 y	49-408 µg/24 h		284.4-2069 nmol/24 h	
		10-13 y	110-714 µg/24 h		557.7-3620 nmol/24 h	
		14-17 y	107-741 µg/24 h		542.5-3757 nmol/24 h	
		18-29 y	94-604 µg/24 h		476.6-3062 nmol/24 h	
		30-39 y	115-695 µg/24 h		583.1-3524 nmol/24 h	
		40-49 y	182-739 µg/24 h		922-3747 nmol/24 h	
		≥50 y	224-832 µg/24 h		1136-4218 nmol/24 h	

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Metanephrines, plasma, free fractionated	Plasma	Metanephrine Normetanephrine Total metanephrines	≤57 pg/mL ≤148 pg/mL ≤205 pg/mL		≤0.29 nmol/L ≤0.75 nmol/L ≤1.04 nmol/L	Maintain on ice until centrifuged. Separate plasma within 2 h of venipuncture. Freeze plasma and store frozen until assay at −20°C. Assay by liquid chromatography, tandem mass spectrometry (LC/MS/MS). Requires minimum of 1.5 mL refrigerated EDTA plasma. Patient should refrain from alcohol, coffee, tea, tobacco, and strenuous exercise. Patient should be seated and relaxed before the sample is collected. Overnight fasting is recommended. Increased in pheochromocytoma, neuroblastoma, and stress.

Metyrapone stimulation test: Metyrapone in a dose of 30 mg/kg is administered orally at midnight. The 8 AM plasma 11-deoxycortisol is >7 μg/dL (>0.2 nmol/L) and plasma ACTH is >100 pg/mL (22 pmol/L). (See Chapter 4.)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Osmolality	Serum Urine	Random specimen	283-301 mmol/kg 300-900 mmol/kg	1.00	283-301 mmol/kg 300-900 mmol/kg	0.2-mL room temperature serum. 1-mL refrigerated urine.
Osteocalcin	Serum	Males and females 5-9 y 10-13 y 14-17 y Males Females Adults Males >17 y Females >17 y	47-142 ng/mL 49-167 ng/mL 26-203 ng/mL 14-85 ng/mL 9-38 ng/mL 8-32 ng/mL	1.0	47-142 ng/mL 49-167 ng/mL 26-203 ng/mL 14-85 ng/mL 9-38 ng/mL 8-32 ng/mL	Overnight fast preferred. Store refrigerated.
Pancreatic polypeptide	Serum	Adults 18-29 y 30-39 y 40-49 y 50-62 y >62 y Children <3 y 3-9 y 10-13 y 14-17 y Cord blood Term infants Preterm infants Term infants (6 d) 1 mo-2 y 3-7 y >7 y	≤480 pg/mL 70-400 pg/mL 70-430 pg/mL 100-780 pg/mL Not established Not established ≤519 pg/mL ≤361 pg/mL ≤297 pg/mL ≤163 pg/mL ≤180 pg/mL ≤276 pg/mL ≤644 pg/mL ≤685 pg/mL 80-270 pg/mL	0.246	≤118.1 nmol/L 17.2-98.4 nmol/L 17.2-105.8 nmol/L 24.6-191.9 nmol/L Not established Not established ≤127.7 nmol/L ≤88.8 nmol/L ≤73.1 nmol/L ≤40.1 nmol/L ≤44.3 nmol/L ≤67.9 nmol/L ≤158.4 nmol/L ≤168.5 nmol/L 19.7-66.4 nmol/L	Collect serum, process within 4 h, and freeze until assay.

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Parathyroid hormone	Serum	Adults	12-65 pg/mL	0.100	1.3-6.8 pmol/L	Intact hormone measured by immunochemiluminometric assay. Freeze serum at −20°C.
Parathyroid hormone-related protein	Plasma				<2.0 pmol/L	Immunochemiluminometric assay. Pre-chill lavender top tube in ice slurry before collection. Deliver sample immediately to the lab on ice for processing. Centrifuge in refrigerated centrifuge and store plasma frozen.
Pregnanetriol	24-h urine	Males Females	71-1000 mg/d 47-790 mg/d	2.97	211-2970 µmol/d 140-1436 µmol/d	Refrigerate urine during the collection and collect without preservatives. Gas chromatography/mass spectrometry (GC/MS) method. Accumulates in urine when 17-hydroxyprogesterone levels are high.
Pregnenolone	Serum			0.0318		Useful in the diagnosis of congenital adrenal hyperplasia and in virilizing adrenal tumors.
		Children Males				
		0-6 y	Not established		Not established	
		7-9 y	<206 ng/dL		<6.55 nmol/L	
		10-12 y	<152 ng/dL		<4.84 nmol/L	
		13-15 y	18-197 ng/dL		0.57-6.26 nmol/L	
		16-17 y	17-228 ng/dL		0.54-7.25 nmol/L	
		Tanner Stages: Stage I	<157 ng/dL		<4.99 nmol/L	
		Stage II	<144 ng/dL		<4.58 nmol/L	
		Stage III	<215 ng/dL		<6.84 nmol/L	
		Stage IV-V	19-201 ng/dL		0.60-6.39 nmol/L	
		Females 0-6 y	Not established		Not established	
		7-9 y	<151 ng/dL		<4.80 nmol/L	
		10-12 y	19-220 ng/dL		0.60-7.00 nmol/L	
		13-15 y	22-210 ng/dL		0.70-6.68 nmol/L	
		16-17 y	22-229 ng/dL		0.70-7.28 nmol/L	
		Tanner Stages: Stage I	<172 ng/dL		<5.47 nmol/L	
		Stage II	22-229 ng/dL		0.70-7.28 nmol/L	
		Stage III	34-215 ng/dL		1.08-6.84 nmol/L	
		Stage IV-V	26-235 ng/dL		0.83-7.47 nmol/L	
		Adults >18 y	33-248 ng/dL		1.05-7.89 nmol/L	

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Progesterone	Serum	Adults Males		0.0318		Chemiluminescent assay. Freeze at −20°C. For children, an ultrasensitive method is used (LC.MS/MS)
		>18 y Females Menstruating females	0.3-1.70 µg/L		0.00954-0.054 nmol/L	
		Follicular phase	0.15-1.40 µg/L		0.00477-0.04452 nmol/L	
		Luteal phase	3.34-25.56 µg/L		0.106-0.813 nmol/L	
		Mid-luteal phase	4.44-28.03 µg/L		0.141-0.891 nmol/L	
		Postmenopausal	<0.73 µg/L		<0.0232 nmol/L	
		Pregnant females				
		First trimester	11.22-90.00 µg/L		0.357-2.862 nmol/L	
		Second trimester	25.55-89.40 µg/L		0.812-2.84 nmol/L	
		Third trimester	48.40-422.50 µg/L		1.539-13.436 nmol/L	
		Results in Ultrasensitive Assay				
		Males				
		5-9 y	≤0.7 ng/mL		≤0.0223 nmol/L	
		10-13 y	≤1.2 ng/mL		≤0.0382 nmol/L	
		14-17 y	≤0.8 ng/mL		≤0.0254 nmol/L	
		18-29 y	≤0.3 ng/mL		≤0.00954 nmol/L	
		30-39 y	≤0.2 ng/mL		≤0.00636 nmol/L	
		40-49 y	≤0.2 ng/mL		≤0.00636 nmol/L	
		50-59 y	≤0.7 ng/mL		≤0.0223 nmol/L	
		Females				
		5-9 y	0.6 ng/mL		0.0191 nmol/L	
		10-13 y	10.2 ng/mL		0.324 nmol/L	
		14-17 y	11.9 ng/mL		0.378 nmol/L	
		Early follicular phase	≤0.6 ng/mL		≤0.0191 nmol/L	
		Late follicular phase	≤2.7 ng/mL		≤0.857 nmol/L	
		Mid-cycle phase	≤16.1 ng/mL		≤0.512 nmol/L	
		Luteal phase	≤3.0-31.4 ng/mL		≤0.0954-0.999 nmol/L	
		Postmenopausal women	≤0.2 ng/mL		≤0.00636 nmol/L	
Proinsulin	Serum	Adults	≤18.8 pmol/L	1.0	≤18.8 pmol/L	Radioimmunoassay on frozen serum. Useful in the diagnosis of insulinoma.
Prolactin	Serum	Adult		0.045		Freeze serum at −20°C. Because of circadian changes in prolactin secretion, with peaks at night and in the afternoon, specimens should be drawn at the same time of day for comparability, preferably after fasting overnight.
		Men >18 y Women	2.1-17.7 µg/L		0.0945-0.797 nmol/L	
		Non-pregnant	2.8-29.2 µg/L ng/mL		0.126-1.31 nmol/L	
		Pregnant	9.7-208.5 µg/L ng/mL		0.436-9.38 nmol/L	
		Postmenopausal	1.8-20.0 µg/L ng/mL		0.081-0.9 nmol/L	
		Children Males				
		Tanner I	≤10 ng/mL		≤0.45 nmol/L	
		Tanner II-III	≤6.1 ng/mL		≤0.3 nmol/L	
		Tanner IV-V	2.8-11.0 ng/mL		0.13-0.5 nmol/L	
		Females				
		Tanner I	3.6-12.0 ng/mL		0.03-0.54 nmol/L	
		Tanner II-III	2.6-18.0 ng/mL		0.117-0.81 nmol/L	
		Tanner IV-V	3.2-20.0 ng/mL		0.144-0.9 nmol/L	

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Renin	Plasma	Normal sodium diet (75-150 mmol/d) Recumbent	0.25-5.82 ng/ mL/h			Discontinue diuretics, estrogens and oral contraceptives for at least 2 wk and non-diuretic anti-hypertensive Rx several days before study. Dietary sodium content should be main-tained for at least 3 d prior to testing. Collection instruc-tions: Supine speci-mens should be drawn in the morning, prefer-ably after over-night recumbency, but in any case, after a minimum of 30 min supine prior to drawing the specimen. Collect in lavender tube and transport to the laboratory at room temperature for prompt separa-tion and freezing of plasma. Do not place on ice or refrigerate sample.
Sex hormone–binding globulin	Serum	1 mo-2 y	60-252 nmol/L	1.0	60-252 nmol/L	Store refrigerated, collect room temperature and process serum sample immedi-ately and store frozen at –20°C. Two-site immuno-radiometric assay.
		1-8 y	72-220 nmol/L		72-220 nmol/L	
		Pubertal males	16-100 nmol/L		16-100 nmol/L	
		Pubertal females	36-125 nmol/L		36-125 nmol/L	
		>18 y old males	20-60 nmol/L		20-60 nmol/L	
		>18 y old females	40-120 nmol/L		40-120 nmol/L	
		Postmenopausal females	28-112 nmol/L		28-112 nmol/L	
Somatostatin	Plasma	Adults	Up to 25 pg/mL	0.426	Up to 10.65 pmol/L	Draw in prechilled tube, separate plasma, and freeze immediately.
Testosterone, total	Serum			0.0347		Freeze at –20°C. Two-step chemilu-minescent microparticle immunoassay.
		Males >18 y	240-871 ng/dL		8.33-30.22 nmol/L	
		Females >18 y	9-55 ng/dL		0.312-1.91 nmol/L	
		Ultrasensitive Assay Pediatric – Males				
		Premature (26-28 wk)	59-125 ng/dL		2.05-4.34 nmol/L	
		Premature (31-35 wk)	37-198 ng/dL		1.28-6.87 nmol/L	Ultrasensitive for pediatric range. LC/MS/MS
		Newborn	37-198 ng/dL		1.28-6.87 nmol/L	
		1-5 mo	14-363 ng/dL		0.49-12.6 nmol/L	
		6-24 mo	<37 ng/dL		<1.28 nmol/L	

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
		2-3 y	<15 ng/dL		<0.52 nmol/L	
		4-5 y	<19 ng/dL		<0.66 nmol/L	
		6-7 y	<13 ng/dL		<0.45 nmol/L	
		8-9 y	2-8 ng/dL		0.069-0.28 nmol/L	
		10-11 y	2-165 ng/dL		0.069-5.73 nmol/L	
		12-13 y	3-619 ng/dL		0.104-21.48 nmol/L	
		14-15 y	31-733 ng/dL		1.08-25.45 nmol/L	
		16-17 y	158-826 ng/dL		5.48-28.66 nmol/L	
		18-39 y	300-1080 ng/dL		10.41-37.48 nmol/L	
		40-59 y	300-890 ng/dL		10.41-30.88 nmol/L	
		60 and older	300-720 ng/dL		10.41-24.98 nmol/L	
		Pediatric – Females Premature (26-28 wk)	5-16 ng/dL		0.17-0.55 nmol/L	
		Premature (31-35 wk)	5-22 ng/dL		0.17-0.76 nmol/L	
		Newborn	20-64 ng/dL		0.69-2.22 nmol/L	
		1-5 mo	<20 ng/dL		<0.69 nmol/L	
		6-24 mo	<9 ng/dL		<0.31 nmol/L	
		2-3 y	<20 ng/dL		<0.69 nmol/L	
		4-5 y	<30 ng/dL		<1.04 nmol/L	
		6-7 y	<7 ng/dL		<0.24 nmol/L	
		8-9 y	1-11 ng/dL		0.035-0.38 nmol/L	
		10-11 y	3-32 ng/dL		0.10-1.11 nmol/L	
		12-13 y	6-50 ng/dL		0.21-1.74 nmol/L	
		14-15 y	6-52 ng/dL		0.21-1.80 nmol/L	
		16-17 y	9-58 ng/dL		0.31-2.01 nmol/L	
		18-39 y	9-55 ng/dL		0.31-1.91 nmol/L	
		40-59 y	9-55 ng/dL		0.31-1.91 nmol/L	
		60 and older	5-32 ng/dL		0.17-1.11 nmol/L	
Testosterone, free	Serum	Male		3.47		Freeze at −20°C. Tracer equilibrium dialysis.
		5-9 y	≤5.3 pg/mL		≤18.4 pmol/L	
		10-13 y	0.7-52.0 pg/mL		2.4-180 pmol/L	
		14-17.9 y	18.0-111 pg/mL		62.5-385 pmol/L	
		Adults				
		18-69 y	35-155 pg/mL		121.4-5379 pmol/L	
		70-89 y	30-135 pg/mL		104.1-468.4 pmol/L	
		Female				
		5-9 y	0.2-5.0 pg/mL		0.69-17.3 pmol/L	
		10-13 y	0.1-7.4 pg/mL		0.35-25.7 pmol/L	
		14-17 y	0.5-3.9 pg/mL		1.73-13.5 pmol/L	
		Adults				
		18-69 y	0.1-6.4 pg/mL		0.35-22.2 pmol/L	
		70-89 y	0.2-3.7 pg/mL		0.69-12.8 pmol/L	

(*Continued*)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Thyroglobulin	Serum	Adult Males Females	1.4-29.2 ug/L 1.5-38.5 ug/L	1.00		Freeze at −20°C. Presence of thyroglobulin autoantibodies in the patient's serum may falsely lower the result. Ultrasensitive assay. Immuno-enzymatic assay for thyroglogulin and antibodies.
		Thyroglobulin antibodies	<2.00 IU/L			
Thyroid antibodies	Serum	Thyroperoxidase antibodies	≤100 IU/mL (negative) >100 IU/mL (positive)			Serum collected and refrigerated. ELISA for thyroperoxidase antibodies. Chemiluminescent microparticle immunoassay for anti-thyroglobulin antibodies.
		Thyroglobulin antibodies	≤2 IU/mL (negative)			
Thyroid-stimulating hormone (TSH) ultrasensitive	Serum or heparinized plasma	Adult Males and nonpregnant females	0.45-4.12 mU/L		0.45-4.12 mU/L	Chemiluminescent microparticle immunoassay. Sensitivity: 0.01 mU/mL. TSH levels decline in first several weeks of life, although they remain elevated in some infants. Free T_4 is helpful in the interpretation of TSH.
		Children Term infants <2 mo 2 mo to <6 mo 6 mo to <14 y 14 y to <18 y	1.12-6.31 mU/L 0.73-4.77 mU/L 0.70-4.17 mU/L 0.47-3.41 mU/L		1.12-6.31 mU/L 0.73-4.77 mU/L 0.70-4.17 mU/L 0.47-3.41 mU/L	
Thyroid-stimulating hormone receptor antibody (TSH-R Ab [stim])	Serum		≤140% of basal activity		–	Assay based on cyclic AMP generation in CHO cells transfected with human TSH receptor gene. Also known as TSI.
Thyroid uptake of radioactive iodine (RAIU)	Activity over thyroid gland	Fractional uptake 2 h 6 h 24 h	4-12% 6-15% 8-30%			Ingestion or administration of iodide will decrease thyroid uptake of RAI.
Thyrotropin-binding inhibitory immunoglobulin (TB-II)	Serum	Adults and children Graves disease	≤16% inhibition 16-100% inhibition	–	–	Sensitivity: 2% inhibition. May be useful in Graves disease, neonatal hypothyroidism, and postpartum thyroid dysfunction.
Thyroxine-binding globulin (TBG)	Serum	Men (>18 y) Women (>18 y) Children 4-6 y 7-8 y 9-10 y 11 y 12 y 13 y 14 y 15 y 16 y 17 y	12.7-25.1 µg/mL 13.5-30.9 µg/mL 14.8-32.9 µg/mL 16.3-30.7 µg/mL 15.8-27.4 µg/mL 15.5-27.4 µg/mL 14.8-26.2 µg/mL 13.8-25.2 µg/mL 12.2-25.2 µg/mL 10.8-23.8 µg/mL 10.0-23.8 µg/mL 8.5-23.1 µg/mL	18.5	235-464.5 nmol/L 249.8-571.7 nmol/L 273.8-608.7 nmol/L 301.6-568 nmol/L 292.3-506.9 nmol/L 286.8-506.9 nmol/L 273.8-484.7 nmol/L 255.3-466.2 nmol/L 225.7-466.2 nmol/L 199.8-440.3 nmol/L 185-440.3 nmol/L 157.2-427.4 nmol/L	Chemiluminescent assay. Quantifies TBG to differentiate TBG abnormalities from thyroid dysfunction.

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
		Males				
		Tanner I	13.5-28.4 µg/mL		249.8-525.4 nmol/L	
		Tanner II	15.1-25.9 µg/mL		279.4-479.2 nmol/L	
		Tanner III	14.0-26.3 µg/mL		259-486.6 nmol/L	
		Tanner IV	13.2-25.0 µg/mL		244.2-462.5 nmol/L	
		Tanner V	12.2-23.7 µg/mL		225.7-438.5 nmol/L	
		Females				
		Tanner I	14.2-28.5 µg/mL		262.7-527.3 nmol/L	
		Tanner II	15.0-23.1 µg/mL		277.5-427.4 nmol/L	
		Tanner III	13.7-23.0 µg/mL		253.4-425.5 nmol/L	
		Tanner IV	12.0-22.8 µg/mL		222-421.8 nmol/L	
		Tanner V	9.1-22.8 µg/mL		168.4-421.8 nmol/L	
Thyroxine (T_4)	Serum	Children		12.87		Refrigerate serum. Fasting preferred. Elevated levels in pregnancy due to increased TBG.
		1-8 y	5.9-11.5 µg/dL		75.9-148 nmol/L	
		9-13 y	4.7-10.4 µg/dL		60.5-133.8 nmol/L	
		14-17 y	5.0-9.8 µg/dL		64.3-126.1 nmol/L	
		Adults	4.8-10.4 µg/dL		61.8-133.8 nmol/L	
		Pregnancy				
		First trimester	6.4-15.2 µg/dL		82.4-195.6 nmol/L	
		Second trimester	7.4-15.2 µg/dL		95.2-195.6 nmol/L	
		Third trimester	7.7-13.8 µg/dL		99.1-189.2 nmol/L	
		All trimesters together	7.0-14.7 µg/dL		90.1-189.2 nmol/L	
Thyroxine, free (FT_4)	Serum or plasma			12.87		Two-step immuno-chemiluminescent microparticle immunoassay. Refrigerate serum.
		Male				
		1-3 d			10-36 pmol/L	
		4-30 d			6-30 pmol/L	
		1-12 mo			10-26 pmol/L	
		1-5 y			11-21 pmol/L	
		6-10 y			11-19 pmol/L	
		>10 y			10-18 pmol/L	
		Female				
		1-3 d			11-25 pmol/L	
		3-40 d			8-25 pmol/L	
		1-12 mo			11-24 pmol/L	
		1-5 y			12-19 pmol/L	
		6-10 y			11-19 pmol/L	
		>10 y			10-18 pmol/L	
		Prematures, 25-30 wk, first week of life	0.5-3.3 ng/dL		6.43-42.47 pmol/L	
		Prematures, 31-36 wk, first week of life	1.3-4.7 ng/dL		16.7-60.49 pmol/L	Equilibrium dialysis, RIA; refrigerated serum
		Cord blood, >37 wk	1.2-2.2 ng/dL		15.44-28.3 pmol/L	
		Birth-4 d	2.2-5.3 ng/dL		28.3-68.2 pmol/L	
		2 wk to 2 y	0.8-2.0 ng/dL		10.3-25.74 pmol/L	
		3-20 y	1.0-2.4 ng/dL		12.87-30.89 pmol/L	
		21-87 y	0.8-2.7 ng/dL		10.3-34.75 pmol/L	
		Pregnancy				
		First trimester	0.9-2.0 ng/dL		11.5-25.74 pmol/L	
		Second trimester	0.8-1.5 ng/dL		10.3-19.3 pmol/L	
		Third trimester	0.8-1.7 ng/dL		10.3-21.88 pmol/L	
Triiodothyronine (T_3), total	Serum			0.0154		Refrigerate serum. Chemiluminescent microparticle immunoassay Used to diagnose and monitor treatment of hyperthyroidism.
		Males				
		0-12 mo			0.9-3.1 nmol/L	
		Females				
		0-12 mo			1.6-3.5 nmol/L	
		Males and Females				
		1-5 y			1.6-3.1 nmol/L	
		6-10 y			1.6-2.8 nmol/L	
		11-14 y			1.0-2.9 nmol/L	
		15-18 y			1.1-2.7 nmol/L	
		>18 y			0.9-2.4 nmol/L	

(Continued)

Test	Source	Ages, Conditions, etc	Conventional Units	Conversion Factor	SI Units	Comments
Free T$_3$ (FT$_3$)	Serum	Adults Children ≤1 y 1-9 y 10-13 y 14-18 y	169-370 pg/dL Not established 337-506 pg/dL 335-480 pg/dL 287-455 pg/dL	15.4	2.6-5.7 pmol/L Not established 5.19-7.79 pmol/L 5.16-7.79 pmol/L 4.42-7.01 pmol/L	Nondialysis; two-step chemilumi-nescent microparticle immunoassay on refrigerated serum.
Reverse T$_3$ (RT$_3$)	Serum	Males Females	0.09-0.28 ng/mL 0.07-0.24 ng/mL	0.0154	0.14-0.431 nmol/L 0.108-0.370 nmol/L	LC/MS/MS Assay may be used to establish the cause of abnormal thyroid function tests as due to nonthyroidal illness.
Vanillylmandelic acid (VMA)	24-h urine	1-12 mo 1-2 y 2-5 y 5-10 y >10 y	<36 mg/g Cr <31 mg/g Cr <17 mg/g Cr <15 mg/g Cr <11 mg/g Cr	0.573	<20.63 mmol VMA/mol Cr <17.76 mmol VMA/mol Cr <9.74 mmol VMA/mol Cr <8.59 mmol VMA/mol Cr <6.30 mmol VMA/mol Cr	10 mL aliquot of urine, HPLC detec-tion; 24-h urine is collected with 25 mL of 6N HCl added. Patients should avoid alco-hol, coffee, tea, tobacco, nicotine, bananas, citrus fruit, and heavy exercise prior to the collection.
Vasoactive intestinal peptide	Plasma		22-42 pg/mL	0.30	6.6-12.6 pmol/L	Frozen EDTA plasma. Pre-chill tube and transport to lab on wet ice after collection. Extrac-tion, RIA.
Vitamin D (25-hydroxy)	Serum		20-50 ng/mL	2.50	50-125 nmol/L	Measures both D$_2$ and D$_3$. Freeze serum in plastic tube at −20°C. Chemiluminescent immunoassay.
Vitamin D (1,25 dihydroxy)	Serum	Adults Children 3-7 y	15-60 pg/mL 27-71 pg/mL	2.50	32.5-167 pmol/L 22.5-135 pmol/L	Liquid chromatogra-phy, tandem mass spectrometry. Freeze serum in plastic tube at −20°C.

[a]Adapted from the *Clinical Laboratories Manual* of the University of California Hospital and Clinics, San Francisco, California, September 7, 2015; and, with permission, from Mayo Clinic Rochester 2016 Interpretive Handbook (June 1 2016) normal values for endocrine tests. The factors used in converting conventional units to SI units were derived in part from the *CRC Handbook of Chemistry and Physics* and conversion tables from other online resources. It is important to emphasize that normal ranges vary among different labo-ratories, and the clinician must know the normal range for the test of interest in the laboratory performing the test.

[b]Semen analysis is discussed in Chapter 12.

BCE, bone collagen equivalent.

Index

Page numbers followed by "f" denote figures; those followed by "t" denote tables.